File: E-21-617

FINAL REPORT

OPTIONAL UPDATING STOCHASTIC APPROXIMATION ALGORITHM STUDY

Principal Investigator

James E. Brown, III

March 31, 1974

Prepared for

National Science Foundation
Engineering Division
Electrical Sciences and Analysis Program
NSF Research Initiation Grant No. GK-32699

SUMMARY

The following report summarizes the major developments of the research activity. The specific details of the research activity are outlined in the following sections of the report:

Chapter I is a summary of the research objective, areas of potential application, and technical background.

Chapter II presents the derivation of the optional updating stochastic approximation algorithm and some of the empirical results on the convergence behavior of the algorithm. It is demonstrated that this algorithm has a faster rate of convergence than does the classical stochastic approximation algorithm.

Chapter III re-examines the philosophical motivation underlying the common stopping rules. Based on a Bayesian viewpoint a new stopping rule is considered. The resulting updating algorithm is shown to have better convergence properties than those considered in Chapter II. The concept of terminating boundaries for requiring an update in less than a predetermined number of observations is discussed.

Chapter IV presents the application of the algorithm to a coding problem. For this particular application the algorithm did not offer any reasonable benefit. In hindsight this was to be expected, since the specific coding technique investigated is known to be optimal when a large number of observations are required and the emperical results of Chapters II and III suggest that the algorithm has little to offer then the number of observations are small.

Chapter V presents the theoretical analysis that has been developed for the modified cosh test when used in conjunction with the stochastic approximation algorithm. Due to the analytical problems involved with analyzing

stopping rules for composite hypothesis testing, these results are not complete. However, they do indicate some gross behavior to be expected from the optional updating algorithm. They are consistent with the emperical results.

ENGINEERING RELEVANCE

Despite the significant theoretical developments in the design of optimum stochastic systems, a basic need of the practical design problem is the development of a system which can be optimized without detailed *a priori* knowledge of its input statistics. Stochastic approximation theory provides one method by which this can be done due to its nonparametric nature and computational simplicity. Unfortunately, the classical stochastic approximation algorithms can be inefficient; they require a large number of iterations in order to obtain a result within a given level of confidence.

The purpose of the research described here is to introduce and analyze optional updating stochastic approximation algorithms for system theory applications. With conventional stochastic approximation algorithms, an infinite amount of data is required for training (or adapting) the system parameters. The new algorithm proposed here will allow one to terminate the training phase of the system automatically after the estimated system parameters fall within pre-specified limits of the optimum value. The rate of convergence is also increased, based on the experimental curves presented here.

Representative areas in which this research has engineering application are pattern recognition, system identification, process control, signal filtering and prediction, and coding theory.

I. INTRODUCTION

A. Purpose and Scope of Research

The purpose of the research is to introduce and analyze optional updating stochastic approximation algorithms for system theory applications. Previous investigations into stochastic approximation theory have neglected the cost of sequential sampling. The algorithm proposed here incorporates the sampling cost via Wald's sequential probability ratio test (SPRT) [1]. The novel feature of this algorithm is the use of one sequential procedure, the SPRT, to determine when to apply a second sequential procedure, stochastic approximation [2]-[5].

With conventional stochastic approximation algorithms an infinite amount of data may be required in training (or adapting) the system parameters. By introducing a stopping rule, the training phase of the system may be stopped after a finite interval, once the estimated system parameters fall within prespecified limits of the optimum value. In addition, adaptation does not take place with each measurement. The algorithm uses a set of measurements, the number being determined by the SPRT, in changing the system parameters. Thus, each adaptation is more precise than with stochastic approximation theory. This is a definite advantage if adaptation "noise" is to be minimized.

Representative areas in which this research has direct engineering application are pattern recognition [6]-[11], system identification [12]-[19], process control [20]-[23], signal filtering and prediction [24]-[29], and coding theory [30]-[32]. However, based on the coding results described in Chapter IV, this last area does not seem too fruitful.

This report compares the stochastic approximation algorithm and the optional stopping algorithms primarily on an experimental basis. Chapter V contains the theoretical results that have been obtained. Due to the difficulty in analyzing stopping rules for composite hypothesis testing, a theoretical analysis of these algorithms is hard to come by.

B. Stochastic Approximation

Stochastic approximation is a method for sequentially estimating an unknown parameter when that parameter can not be measured directly. The unknown parameter, θ, is taken to be the unique root of some regression function, M(s). The function M(s) cannot be observed without some measurement noise, Z. The algorithm for generating the sequence of estimates, $\hat{\theta}_n$, for θ is given by

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \mu_n[M(\hat{\theta}_n) + Z_n], \tag{1.1}$$

where μ_n is a parameter which governs the rate of convergence and stability of the algorithm. (A survey of stochastic approximation may be found in [5].)

In one engineering application of stochastic approximation theory, it is desired, once a reasonable system has been built, to optimize its performance by experimenting with the values of its adjustable parameters $s_1, s_2, \ldots, s_p$. The unknown parameter θ denotes the desired values of the above p real numbers; i.e.,

$$\theta^T = (\theta_1, \theta_2, \ldots, \theta_p)^*$$

where

$$s_i = \theta_i \qquad i = 1, 2, \ldots, p$$

denotes the optimum system. Characteristically, the performance of the system is measured by a quadratic loss function in the parameter

---

*T denotes the vector transpose operator.

$s^T = (s_1, s_2, \ldots, s_p)$ of the form

$$L(s) = L_o + \tfrac{1}{2}(s-\theta)^T R(s-\theta),$$

where

L_o = minimum loss possible

and

R = a real, symmetric, positive definite matrix.

The function M(s) is the gradient of L(s) with respect to s; i.e.,

$$M(s) = R(s-\theta).$$

(Since A is positive definite, $s = \theta$ is the unique root of the equation $M(s) = 0$.)

Associated with the cost of using the estimate $\hat{\theta}_n$ is the cost of updating the estimate $\hat{\theta}_n$. This loss is dependent on the cost of estimating $M(\hat{\theta}_n)$ and the cost of performing the operations required by (1.1). This additional cost, C, is independent of $\hat{\theta}_n$ and the time n. When the estimate $\hat{\theta}_n$ is sufficiently close to the parameter θ so that excess cost $\frac{1}{2}(\hat{\theta}_n - \theta)^T R(\hat{\theta}_n - \theta)$ is less than C, it is no longer profitable to update $\hat{\theta}_n$.

Without knowledge of both R and θ, it is impossible to determine when $\hat{\theta}_n$ should no longer be updated by (1.1). However, it should be possible to estimate whether $\hat{\theta}_n$ should be changed or not. A method for achieving this is presented in the next section.

C. Wald's Sequential Probability Ratio Test (SPRT)

Let $Y_1, Y_2, \ldots, Y_n, \ldots$, be a sequence of independent identically distributed random variables according to the probability density function $p_Y(y|\gamma)$, where γ is the parameter to be tested. The problem is to test the hypothesis

$$H_1 : \gamma = \gamma_1$$

against the hypothesis

$$H_o : \gamma = \gamma_o$$

and decide in favor of either H_1 or H_o on the basis of the observations. If H_1 is true, one is to decide in favor of H_1 with probability at least $1 - \beta$, while if H_o is true, one is to decide in favor of H_o with probability at least $1 - \alpha$.

Wald's SPRT is as follows: Continue taking observations as long as

$$B < \lambda_n < A,$$

stop taking observations and decide to accept H_1 as soon as

$$\lambda_n \geq A,$$

and stop taking observations and decide to accept H_o as soon as

$$\lambda_n \leq B,$$

where

$$\lambda_n = \prod_{i=1}^{n} \frac{p_Y(y_i|\gamma_1)}{p_Y(y_i|\gamma_o)} .$$

The constants A and B are related to the error probabilities by

$$A = \frac{1 - \beta}{\alpha}$$

and

$$B = \frac{\beta}{1-\alpha} .$$

The problem of interest in this research is to chose between the composite hypotheses

$$H_1 : 1/2\ (\hat{\theta} - \theta)^T\ R(\hat{\theta} - \theta) > C$$

and

$$H_o : 1/2\ (\hat{\theta} - \theta)^T\ R(\hat{\theta} - \theta) \leq C$$

where the observed random variables are the gradient estimates

$$Y_n = M(\hat{\theta}) + Z_n \quad ,$$

where Z_n is normally distributed with zero mean and variance σ^2. Note that this hypothesis testing problem is equivalent to testing

$$H_1 : \left| \frac{\gamma}{\sigma} \right| \geq \delta$$

against

$$H_o : \left| \frac{\gamma}{\sigma} \right| < \delta \quad ,$$

where the observations are taken from a normally distributed population with mean

$$\gamma = R(\theta - \hat{\theta})$$

and variance σ^2. The tolerance δ is given by

$$\delta = \frac{\sqrt{2Rc}}{\sigma} \quad .$$

While the SPRT is designed for only simple binary hypotheses testing problems it can often be used successfully in a composite binary hypotheses testing problem. Wald [1] suggests the following "cosh" test procedure: Define

$$S_n = \frac{\delta}{\sigma} \sum_{i=1}^{n} Y_i \quad .$$

Continue taking observations as long as

$$\ell n\ B + n\frac{\delta^2}{2} < \ell n \cosh S_n < \ell n\ A + n\frac{\delta^2}{2} \quad ,$$

stop taking observations and decide to accept H_1 as soon as

$$\ell n \cosh S_n > \ell n\ A + n\frac{\delta^2}{2} \quad ,$$

and stop taking observations and decide to accept H_o as soon as

$$\ell n \cosh S_n < \ell n\ B + n\frac{\delta^2}{2} \quad .$$

The test procedure is shown in Figure 1.

The computation of $\ell n \cosh S_n$ is a very cumbersome process. However, as shown by Wald [1], the test is equivalent to continue taking observations as long as

$$b_n < |S_n| < a_n \quad ,$$

stop taking observations and decide to accept H_1 as soon as

$$|S_n| \geq a_n \quad ,$$

and stop taking observations and decide to accept H_o as soon as

$$|S_n| \leq b_n \ ,$$

where

$$a_n = \phi(\ell n\ A + n\ \delta^2/2)$$

$$b_n = \phi(\ell n\ B + n\ \delta^2/2)$$

and

$$\phi(v) = \ell n\left[e^v + \sqrt{e^{2v} - 1}\right] \quad .$$

The sequences $\{a_k\}$ and $\{b_k\}$ can be predetermined before experimentation begins.

If $|S_n| > 3$ when the test is to terminate, an even simpler rule is available: Continue taking observations as long as

$$\ell n\ 2B + n\frac{\delta^2}{2} < |S_n| < \ell n\ 2A + n\frac{\delta^2}{2} \quad ,$$

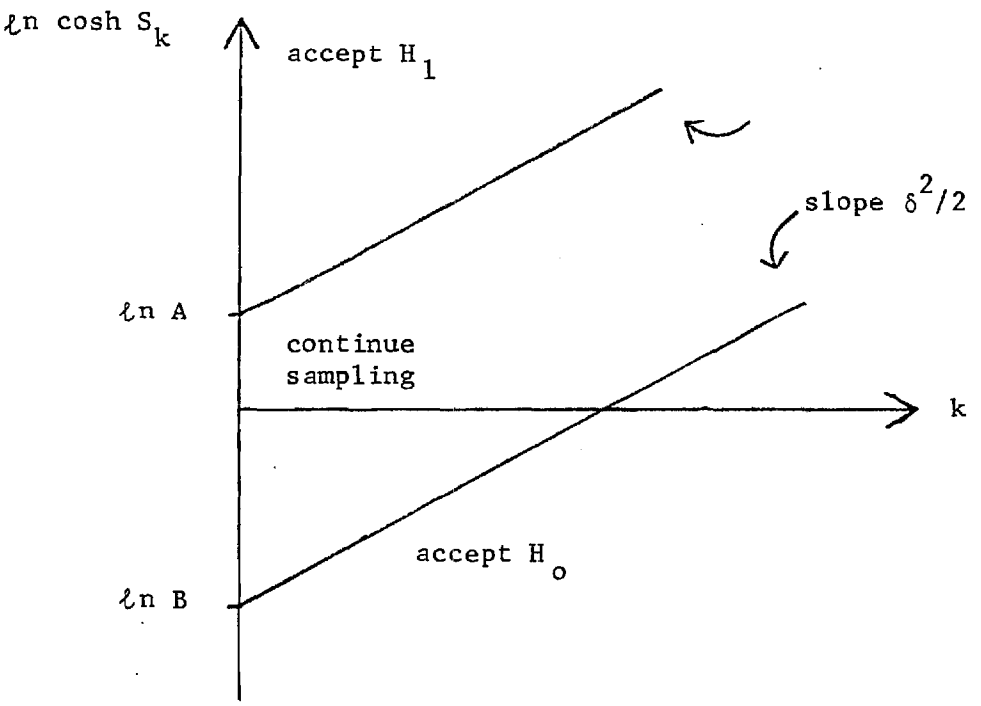


Figure 1

The Test Procedure for the SPRT.

stop and decide to accept H_1 when

$$|S_n| \geq \ell n\ 2A + n\frac{\delta^2}{2} \quad ,$$

and stop and decide to accept H_o when

$$|S_n| \leq \ell n\ 2B + n\frac{\delta^2}{2} \quad .$$

This test does not involve the computation of $\phi(\cdot)$.

Still another version of the SPRT is available. The test consists of paring the observations to obtain the test statistic

$$T_{2n} = \frac{\delta^2}{2\sigma^2} \sum_{i=1}^{n} (Y_{2i-1}^2 + Y_{2i}^2) \quad .$$

Note that

$$X_i = \frac{1}{\sigma} (Y_{2i-1}^2 + Y_{2i}^2)^{\frac{1}{2}}$$

has the Rician probability density function

$$p_{x_i}(x_i) = x_i \exp\left[-\frac{x_i^2 + a^2}{2}\right] I_o(ax_i) \ , \ x_i \geq 0,$$

where

$$a = \sqrt{2}\,\gamma/\sigma \quad .$$

As shown by DiFranco and Rubin [40], the SPRT is given by:

Continue taking observations as long as

$$\ell n\ B + n\ \delta^2\left(1 + \frac{\delta^2}{2}\right) < T_{2n} < \ell n\ A + n\delta^2\left(1 + \frac{\delta^2}{2}\right) \ ,$$

stop and decide H_1 when

$$T_{2n} \geq \ell n\ A + n\delta^2\left(1 + \frac{\delta^2}{2}\right) \ ,$$

and top and decide H_o when

$$T_{2n} \leq \ell n\ B + n\delta^2\left(1 + \frac{\delta^2}{2}\right) \quad .$$

This test is based on the assumption that $\delta \ll 1$. The operating characteristic function (OCF) $\mathcal{L}(\gamma)$ and the average sample number (ASN) are shown in Figures 2 and 3 for various values of α and β.

In passing, it should be pointed out that the problem of determining a stopping time for the stochastic approximation algorithm is equivalent to the determination of confidence intervals for the estimate $\hat{\theta}_n$. By Chebyshev's inequality, one has

$$P_r \left\{ |\hat{\theta}_n - \theta| \geq \epsilon \right\} \leq \frac{b_n^2}{\epsilon^2} .$$

Hence, establishing a stopping rule in terms of b_n^2 will establish a stopping rule in terms of $P_r \left\{ |\hat{\theta}_n - \theta| < \epsilon \right\}$. For example, if the algorithm is stopped when

$$b_n^2 \leq c ,$$

then

$$P_r \left\{ \hat{\theta}_n - \epsilon < \theta < \hat{\theta}_n + \epsilon \right\} \geq 1 - c/\epsilon^2 .$$

The confidence interval is $\left(\hat{\theta}_n - \epsilon, \hat{\theta}_n + \epsilon \right)$ with confidence level $1 - c/\epsilon^2$.

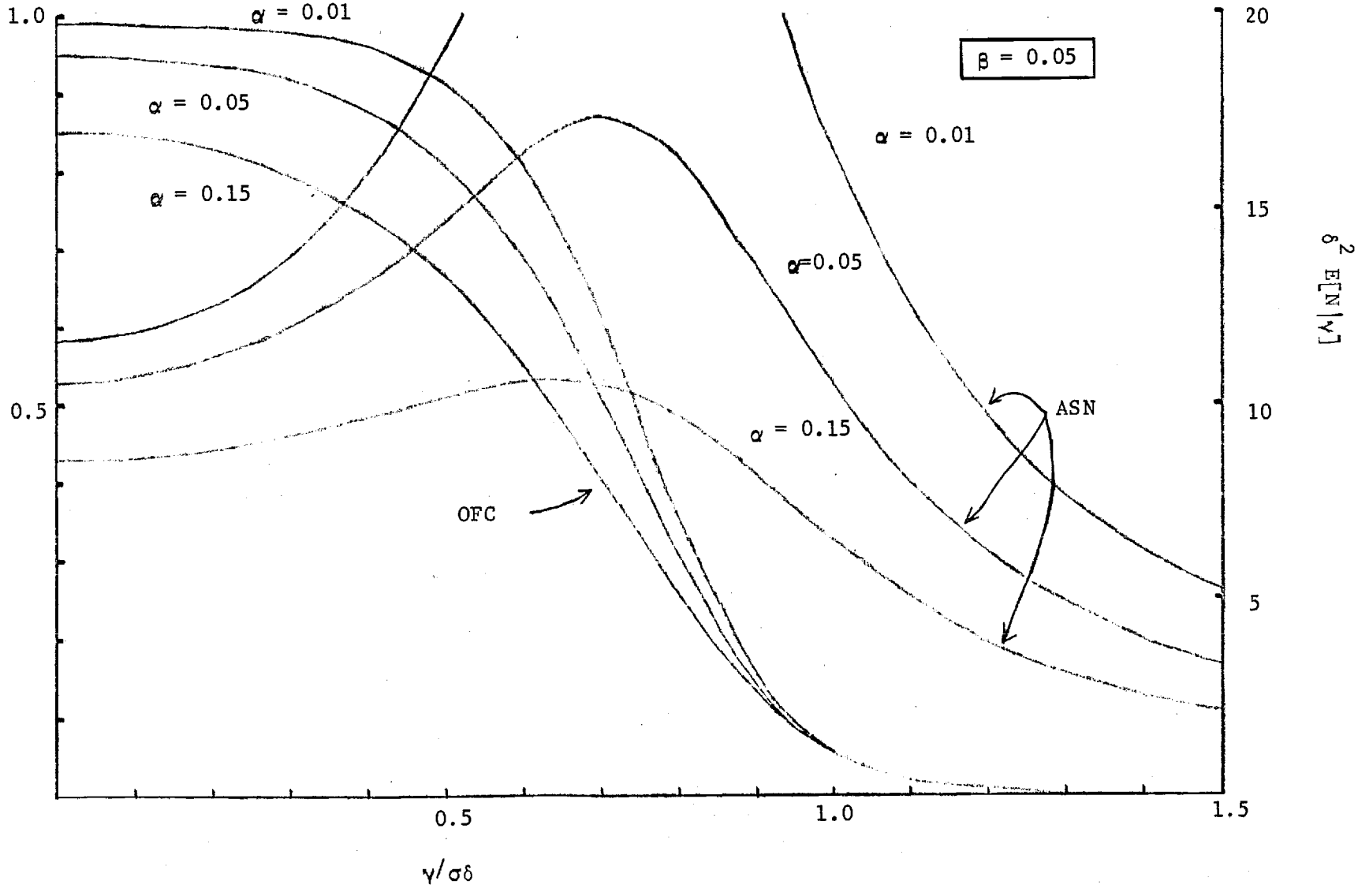


Figure 2. OCF and ASN Curves for Rician Test

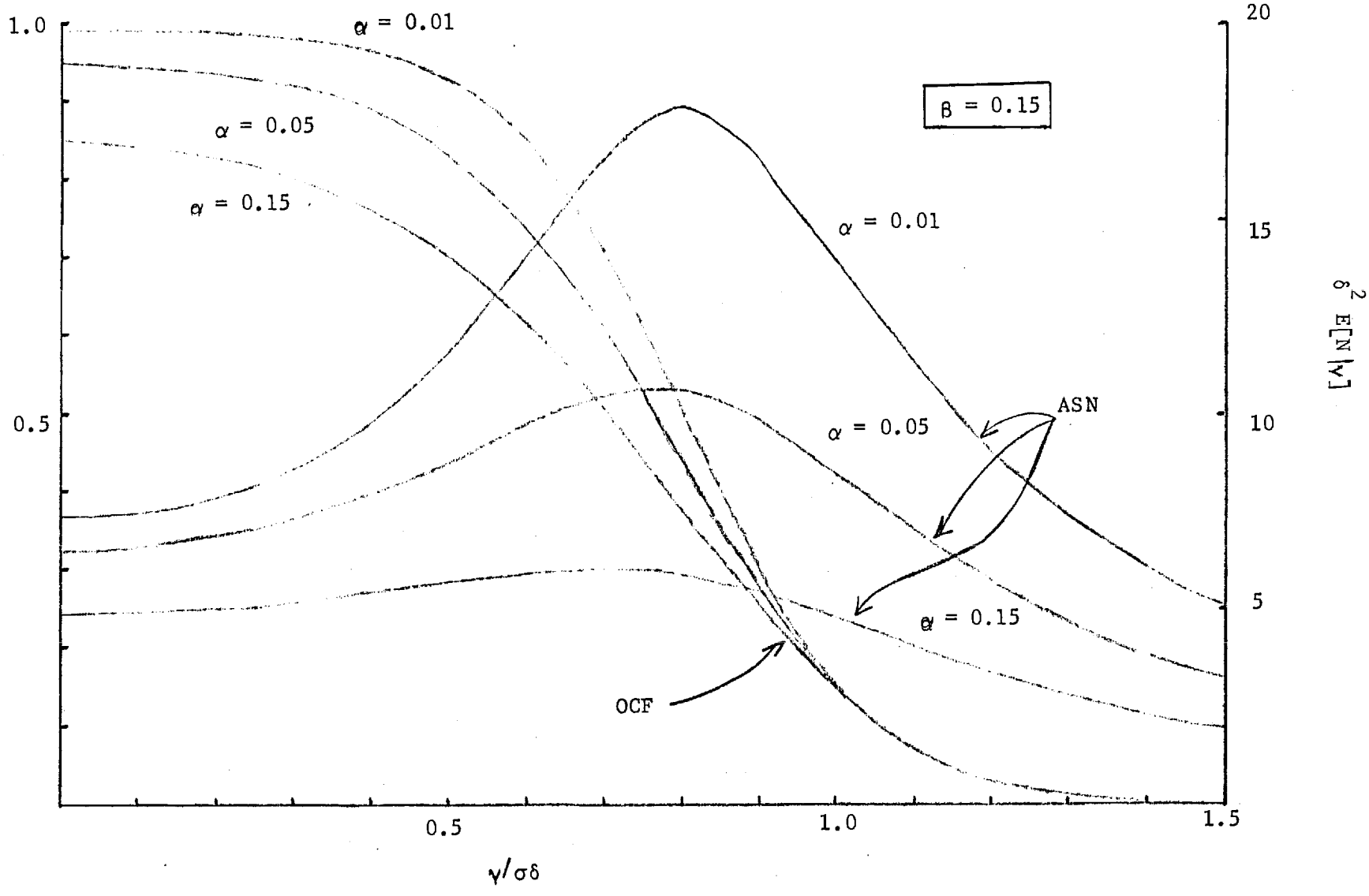


Figure 3. OCF and ASN Curves for Rician Test

II. OPTIONAL UPDATING STOCHASTIC APPROXIMATION ALGORITHMS

A. An Optional Updating Stochastic Approximation Algorithm with Termination

Motivated by the discussion of the SPRT, update the estimate $\hat{\theta}_n$ as follows: Let $\left\{Y_j^{(n)}\right\}$ be the estimates of $M(\hat{\theta}_n)$; i.e.,

$$Y_j^{(n)} = M(\hat{\theta}_n) + Z_j^{(n)},$$

where $\left\{Z_j^{(n)} : 1 \le j < \infty,\ 1 \le n < \infty\right\}$ is the independent, zero-mean, Gaussian measurement process.

Let

$$S_k^{(n)} = \sum_{j=1}^{n} Y_j^{(n)} = k\, M(\hat{\theta}_n) + \sum_{j=1}^{k} Z_j^{(n)} \quad .$$

Let N_n denote the length of time that the estimate $\hat{\theta}_n$ is to be used. The time N_n corresponds to the first k for which

$$\left| S_k^{(n)} \right| \ge \frac{\sigma}{\delta} (\ell n\, A + \ell n\, 2) + k\, \frac{\sigma\delta}{2} .$$

(Pictorally, the update time N_n is determined as shown in Figure 4.)

The estimate $\hat{\theta}_n$ is changed accordingly as

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \mu \cdot \frac{1}{N_n} S_{N_n}^{(n)}. \tag{2.1}$$

Note that

$$\frac{1}{N_n} S_{N_n}^{(n)} = M(\hat{\theta}_n) + \frac{1}{N_n} \sum_{j=1}^{N_n} Z_j^{(n)}.$$

Thus, the algorithm (2.1) has the effect of averaging the gradient estimates and, consequently reducing the variance of the measurement noise. In addition, if for some k,

$$\left| S_k^{(n)} \right| \le \frac{\sigma}{\delta} (\ell n\, B + \ell n\, 2) + k\, \frac{\sigma\delta}{2} ,$$

then it is decided that no further refinement of the estimates for θ is needed and the taking of gradient measurements ceases. The design phase for the system is finished.

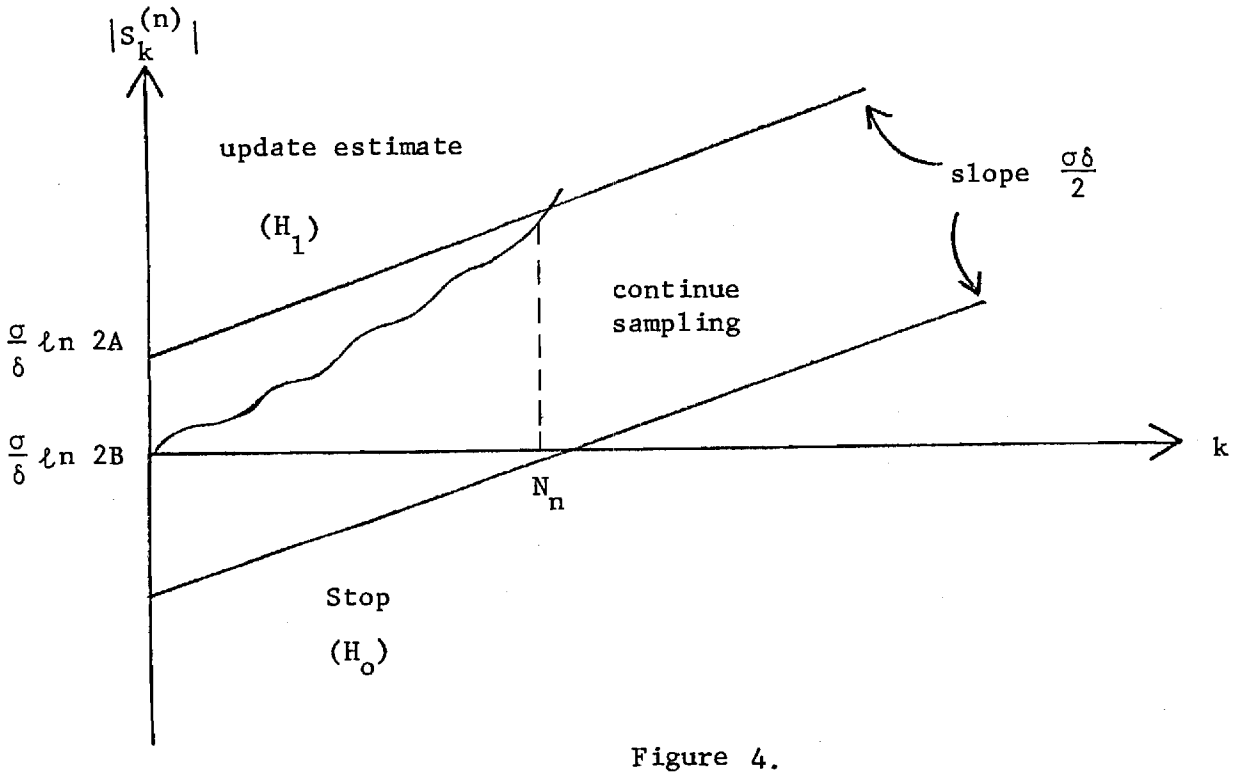


Figure 4.

A Representative Optional Updating Time N_n

Figures 5-13 show a comparison of the experimentally measured mean-square error $E[\|\hat{\theta}_n - \theta\|^2]$ for the optional updating algorithms as a function of the number of gradient measurements. (The abbreviations are defined as EXT = cosh test, SUB = modified cosh test, and MOD = Rician test.) The appropriate value of μ, α, β, R, σ, and the threshold are shown in each figure.

The curves corresponding to c = 0.25 illustrate an important aspect of this algorithm (Figures 7-13). In terms of the rate of convergence and final mean-square error, the following ranking of these simulations is possible as a function of α and β:

α	β	
0.5	0.01	
0.5	0.05	
0.15	0.05	decreasing order
0.10	0.05	of performance
0.05	0.05	
0.05	0.10	
0.15	0.15	

Recalling that the error probabilities correspond to

$$\alpha = P_r \{\text{update made} \mid \text{no update needed}\}$$

$$\beta = P_r \{\text{stop} \mid \text{update needed}\}$$

we see that by allowing more frequent adaptations (large α) than actually required while maintaining close control over stopping errors (small β), the overall performance can be made very satisfactory. This is comforting because the type of error that one wants to control is the error due to stopping while additional adaptations are required. This is controlled

by the value of β. The choice of α is not nearly so critical. Additional adaptations, though not required, are not as serious as stopping too soon. (Of course, choosing α too large defeats the purpose of the algorithm.)

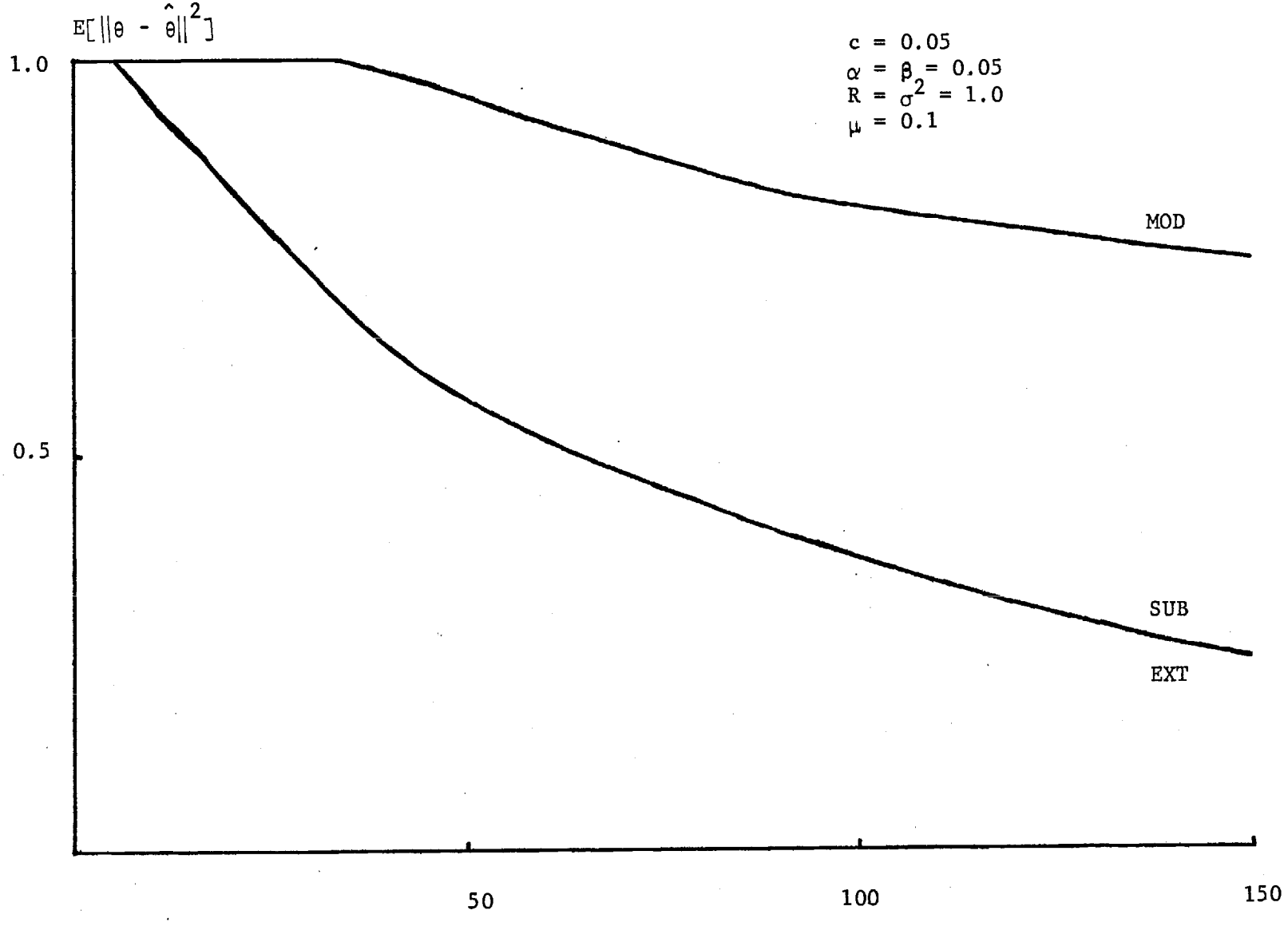


Figure 5. Comparison of Stopping Algorithms.

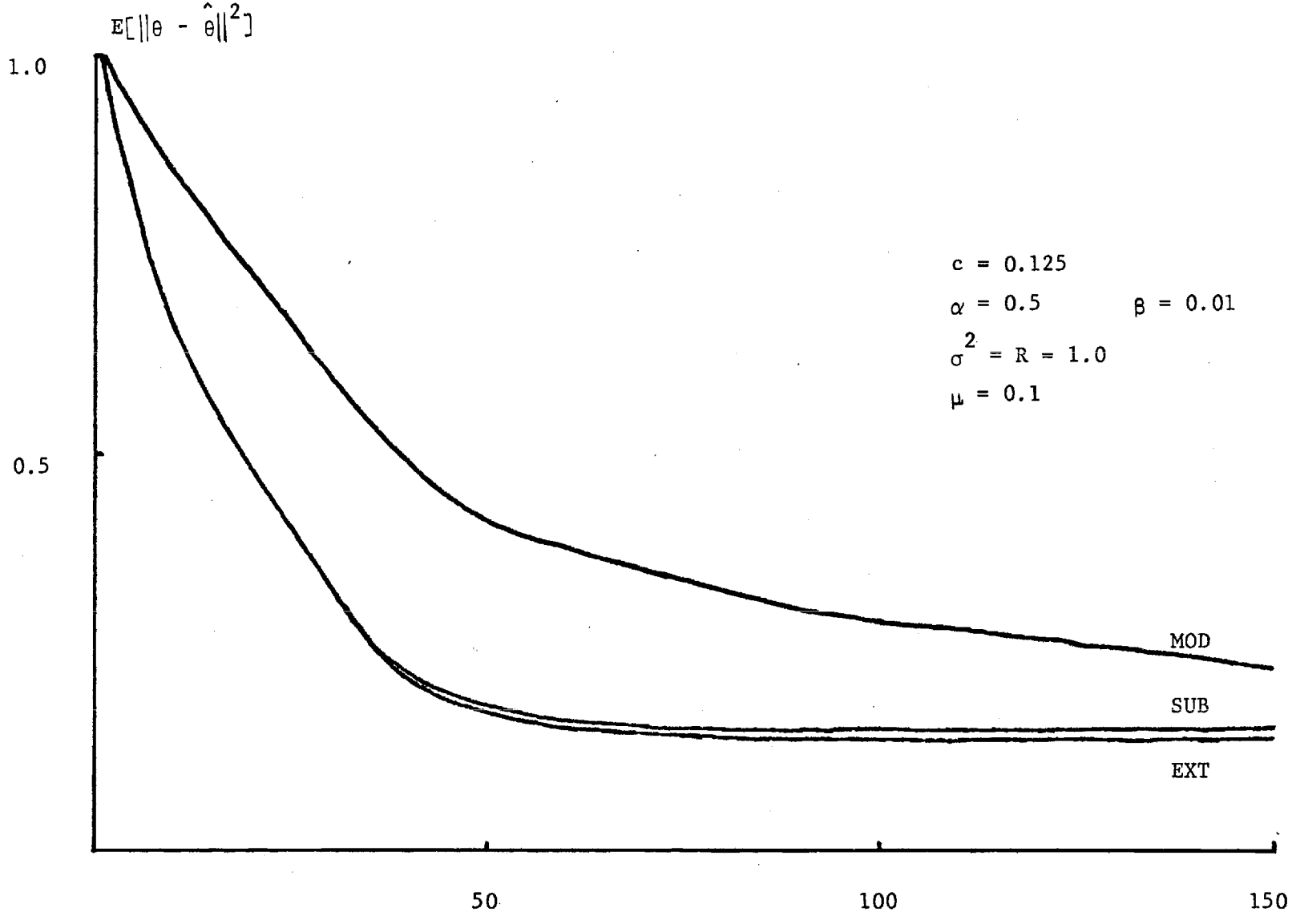


Figure 6. Comparison of Stopping Algorithms.

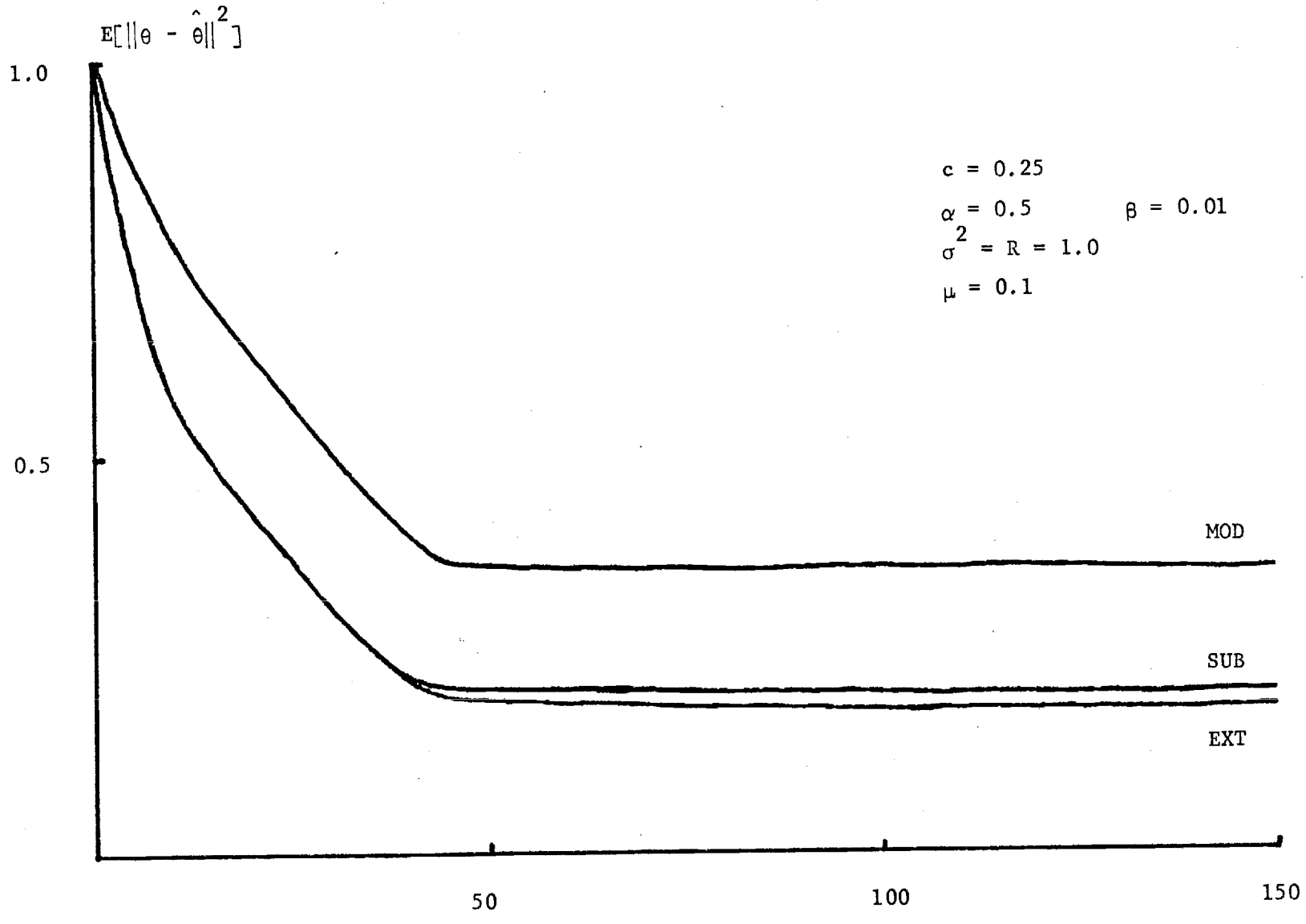


Figure 7. Comparison of Stopping Algorithms.

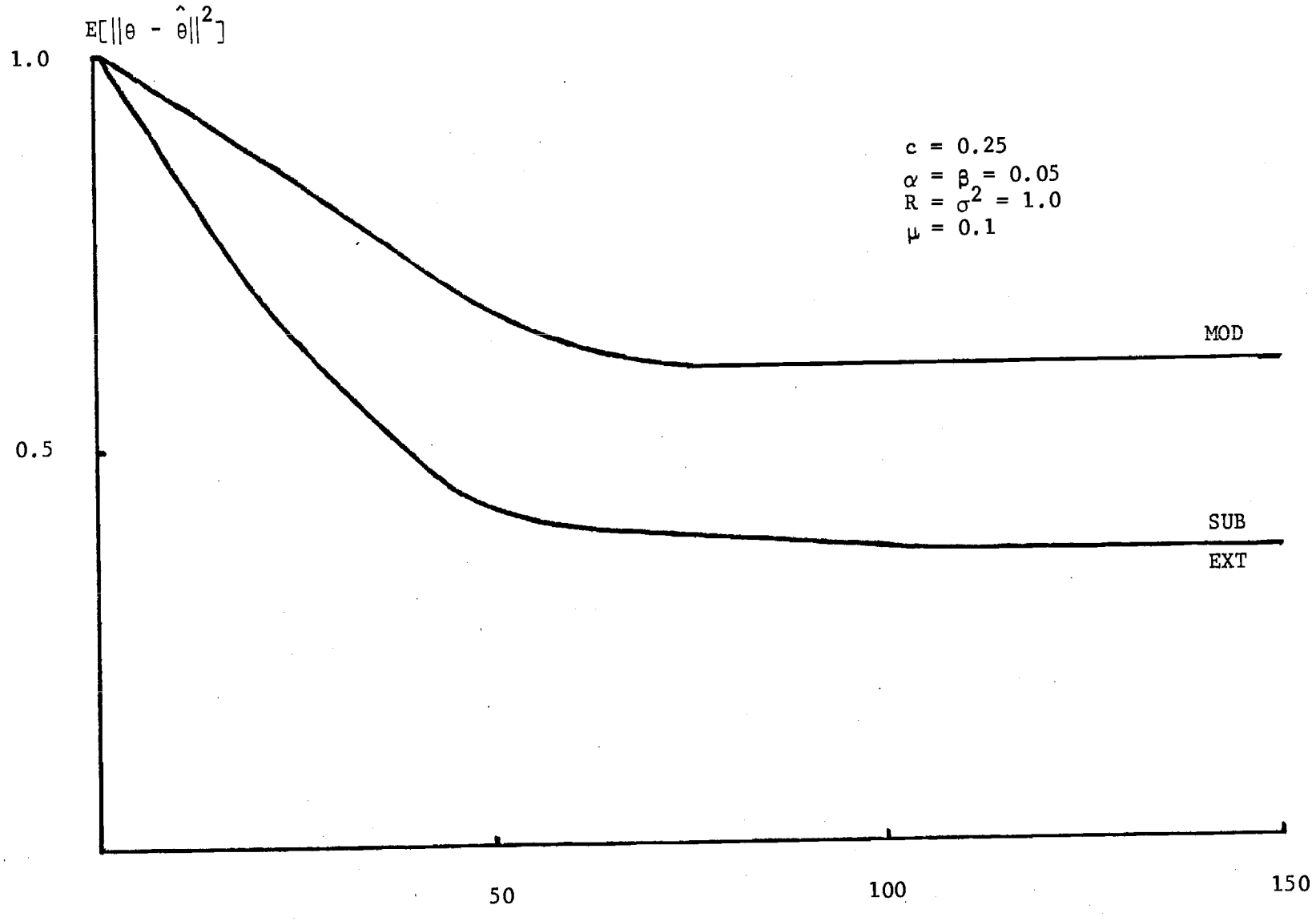


Figure 8. Comparison of Stopping Algorithms.

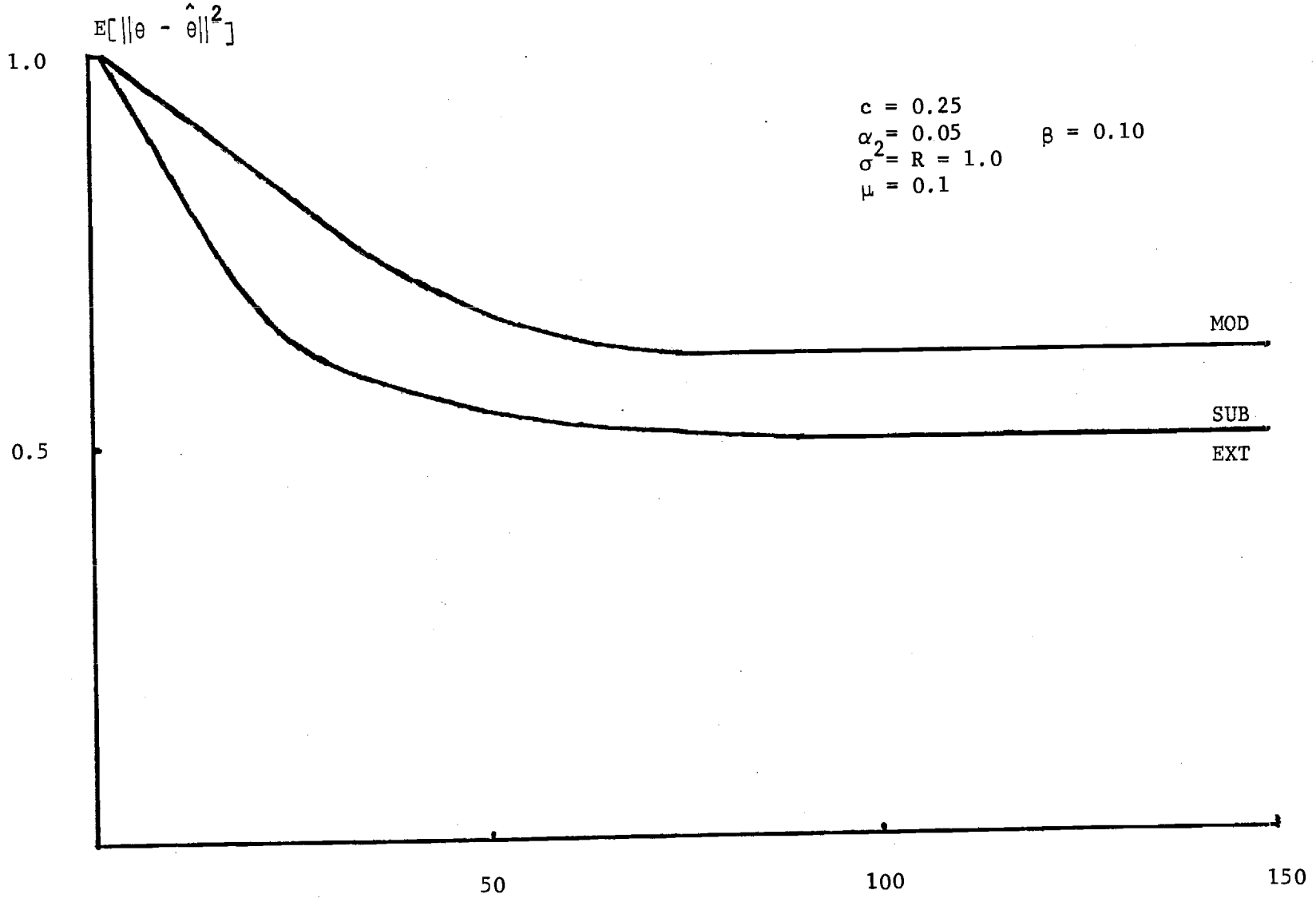


Figure 9. Comparison of Stopping Algorithms.

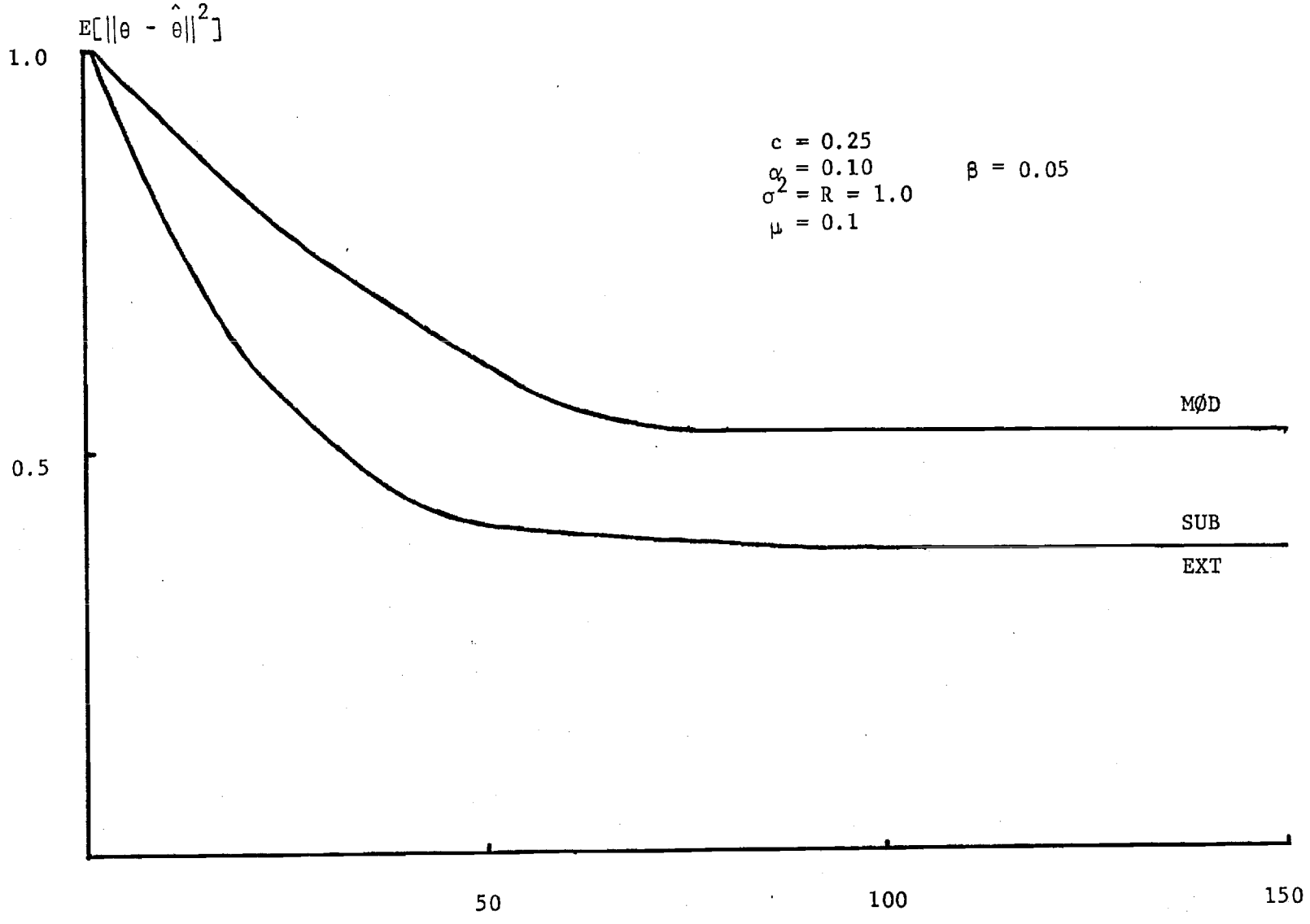


Figure 10. Comparison of Stopping Algorithms.

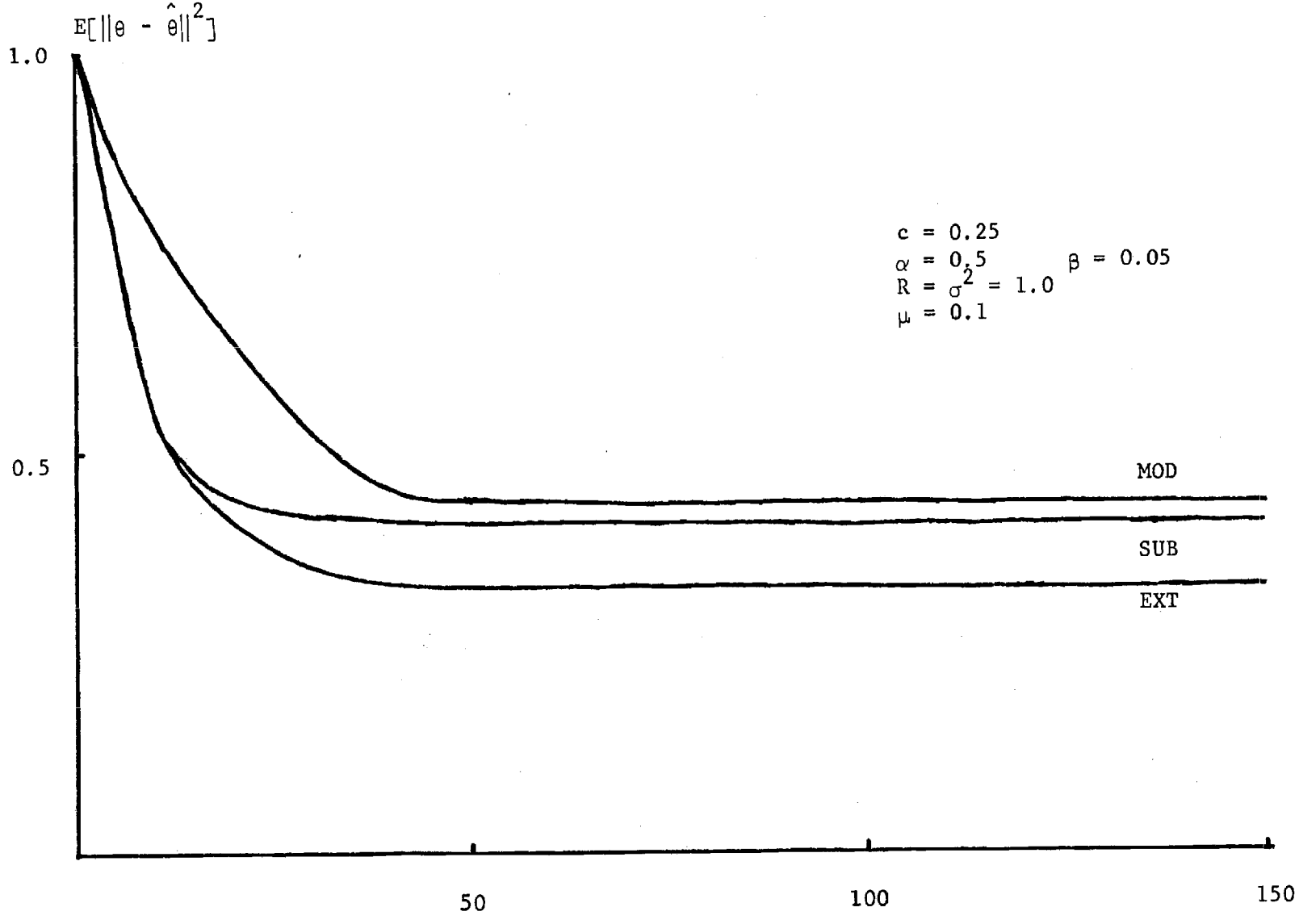


Figure 11. Comparison of Stopping Algorithms.

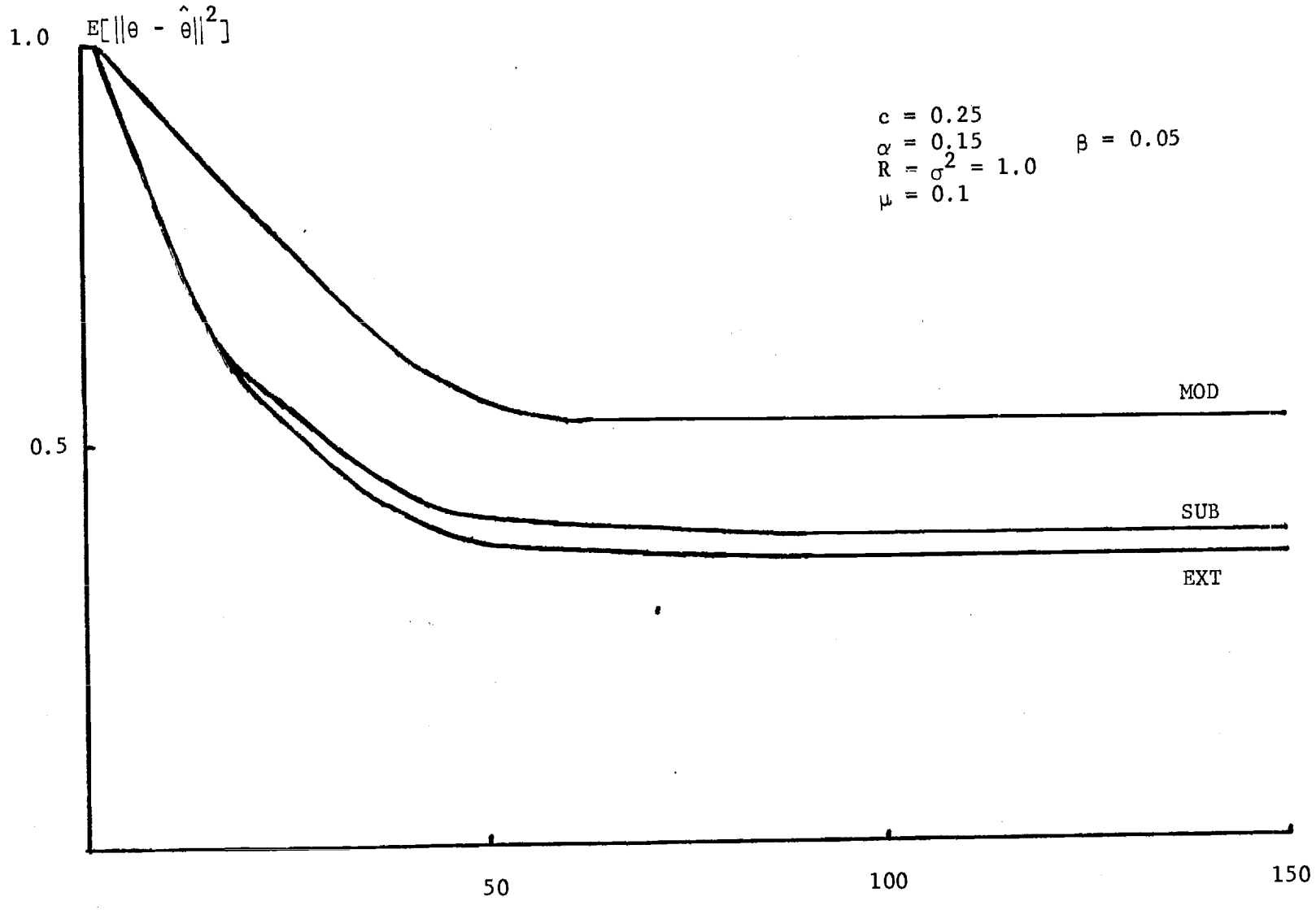


Figure 12. Comparison of Stopping Algorithms.

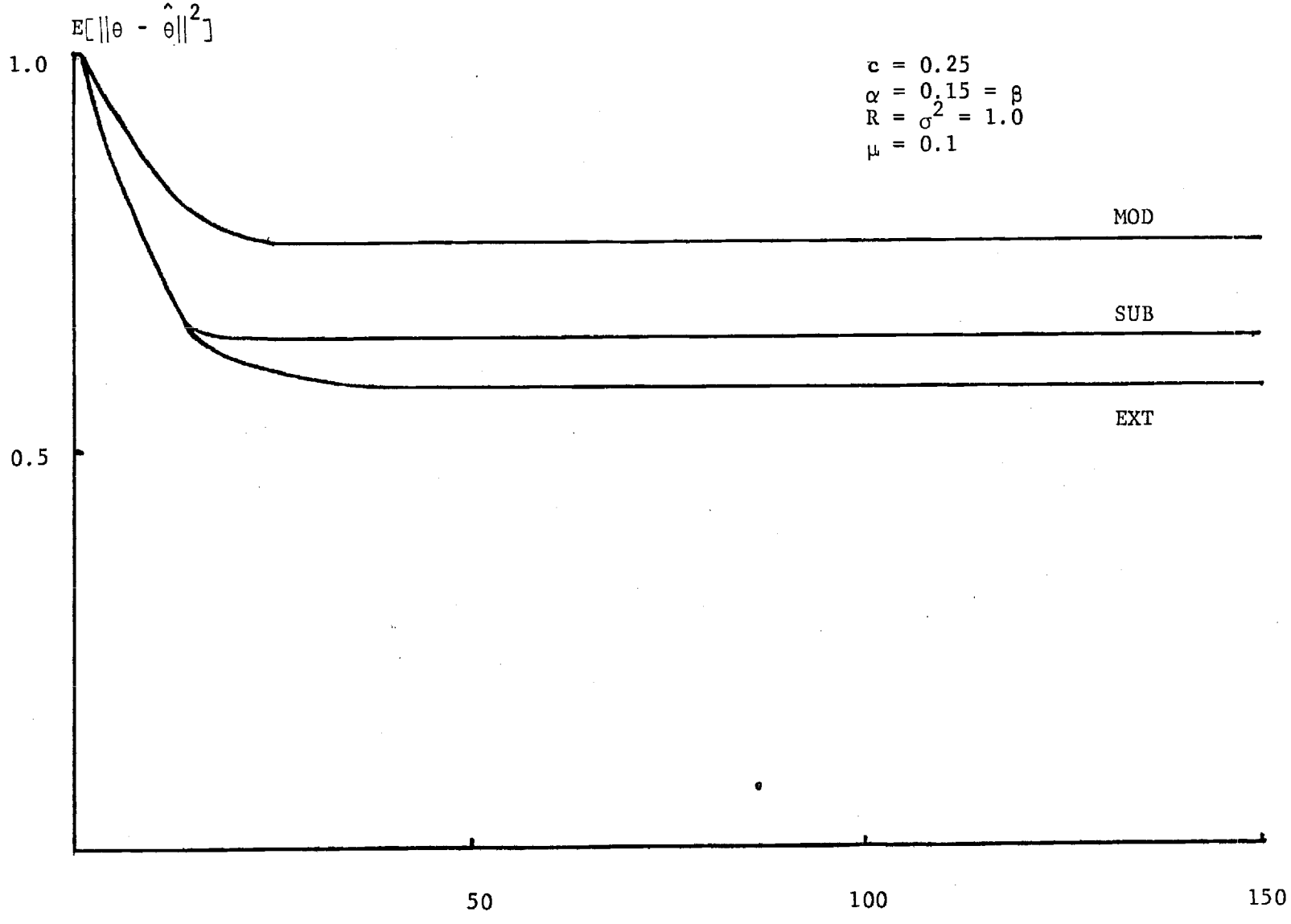


Figure 13. Comparison of Stopping Algorithms.

B. An Optional Updating Stochastic Approximation Algorithm without Termination

In a mildly nonstationary environment in which the unknown parameter θ varies slowly with time, termination of the design phase of the system is undesirable. The parameter θ may drift sufficiently that the system no longer behaves satisfactorily. For this situation, after $|S_k^{(n)}|$ crosses the lower stopping boundary, one would set $S^{(n)} = 0$ and repeat accumulating a new sum $S_k^{(n)}$ until a stopping boundary is crossed. At this time, if $|S_k^{(n)}|$ crosses the upper boundary, the estimate $\hat{\theta}_n$ would be changed according to (2.1); if $|S_k^{(n)}|$ crosses the lower boundary, $S^{(n)}$ would again be set equal to zero and the process would repeat.

If the parameter θ were known to be constant, a similar procedure to the one described above could also be used except when $|S_k^{(n)}|$ crosses the lower boundary, the value of c would be decreased. Computer simulations of this situation are shown in Figures 14-18.* When the decision is made to stop, the value of c is halved and the learning process continues. As can be seen from the figures, the rate of convergence is faster than with the stochastic approximation algorithm. Considerably fewer adaptations are made. It is interesting to note that $\alpha = \beta = 0.15$ results in the best performance in sharp contrast to that when the stopping option is used.

---

*The results of twenty simulations were made for each of several initial values for c. The labeling is the same as in the previous figures with SA corresponding to stochastic approximation.

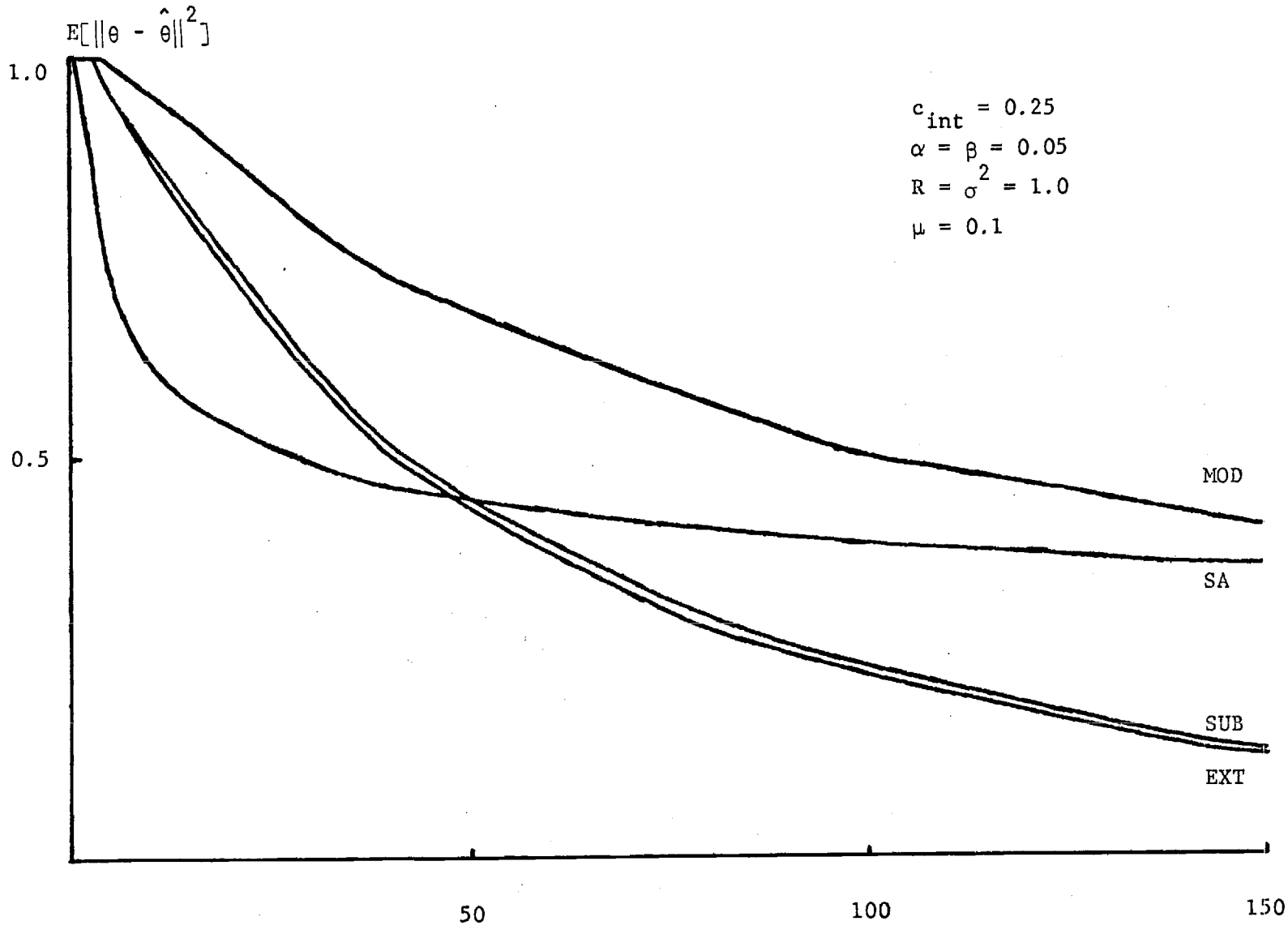


Figure 14. Comparison of Updating Algorithms.

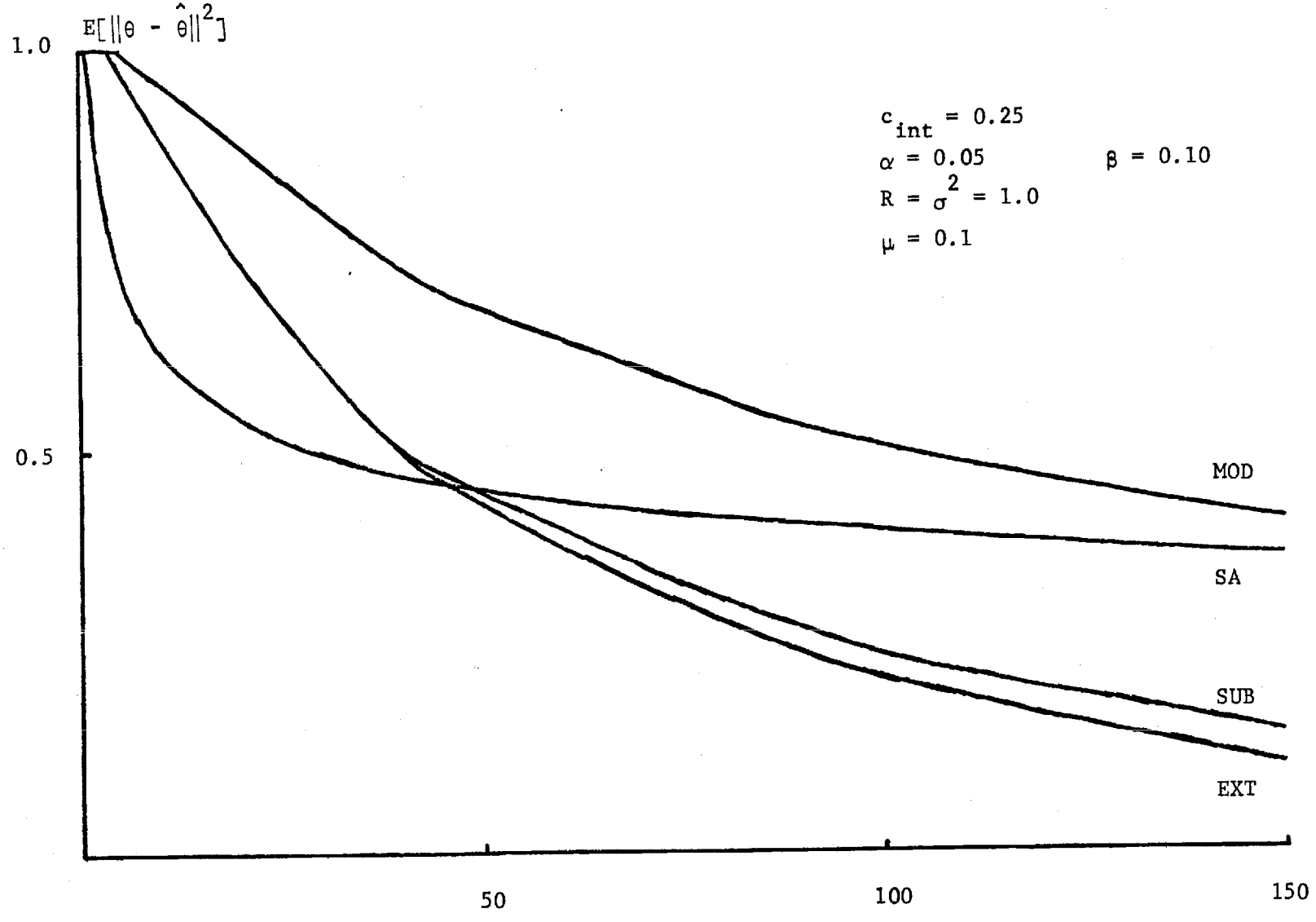


Figure 15. Comparison of Updating Algorithms.

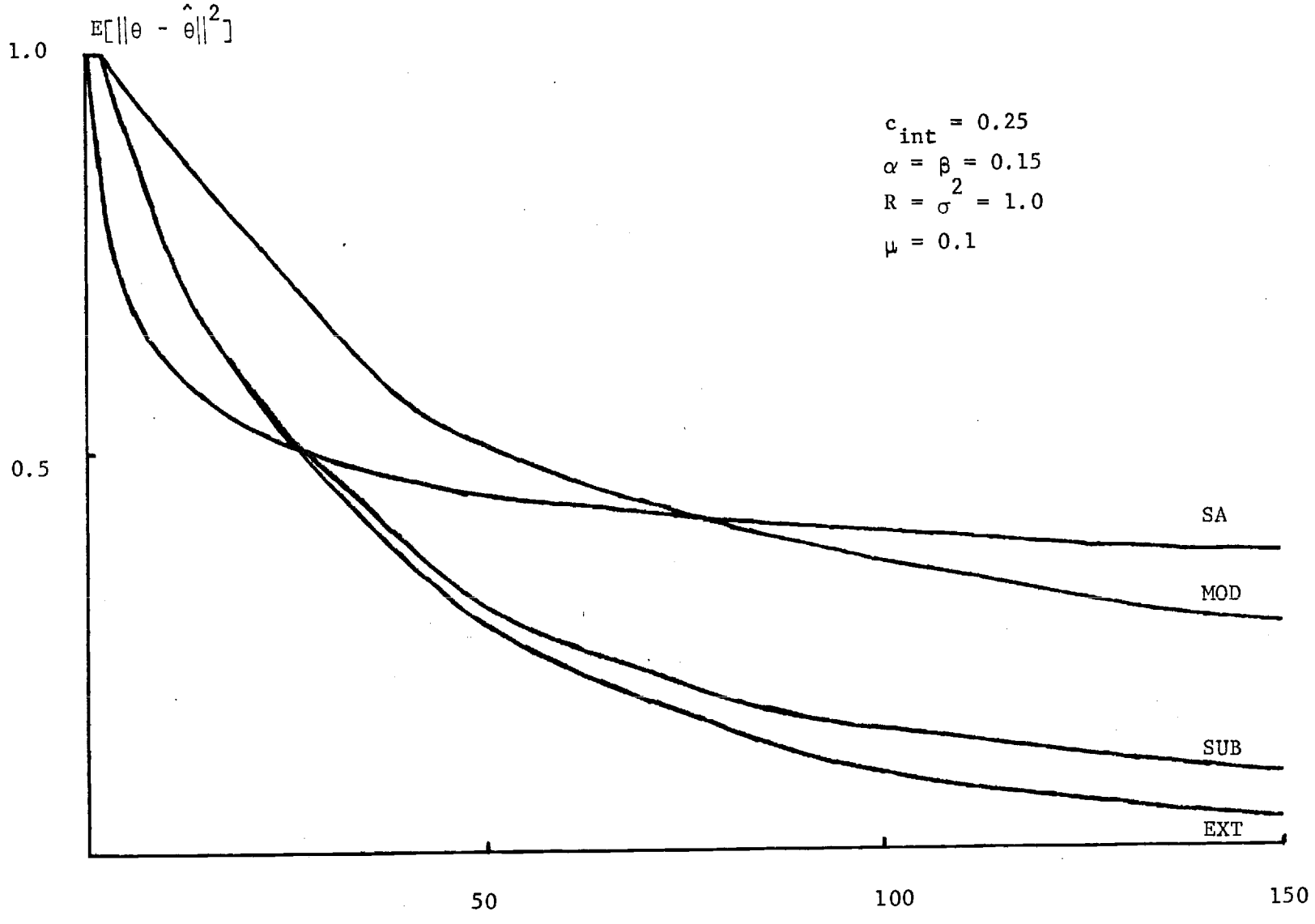


Figure 16. Comparison of Updating Algorithms.

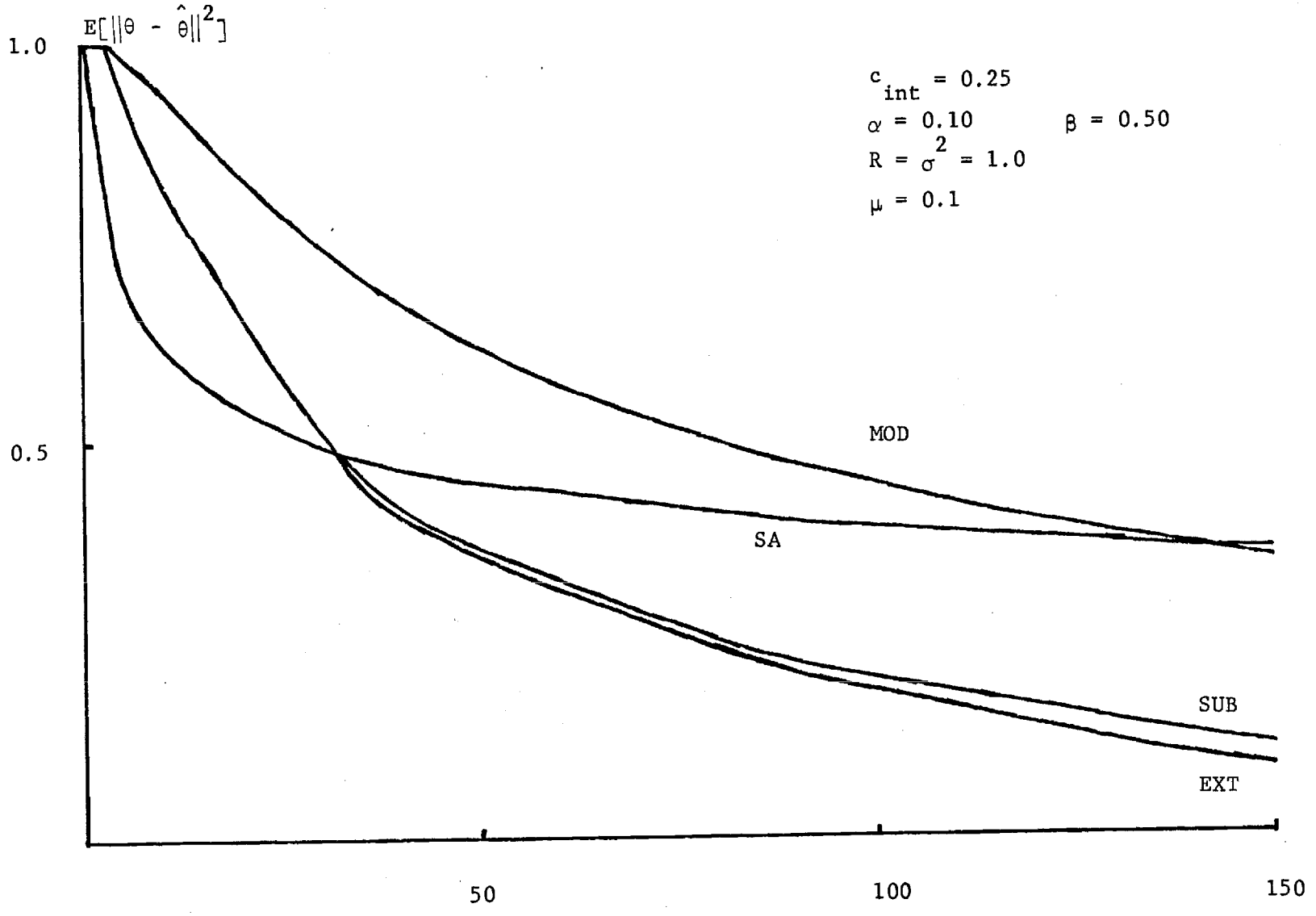


Figure 17. Comparison of Updating Algorithms.

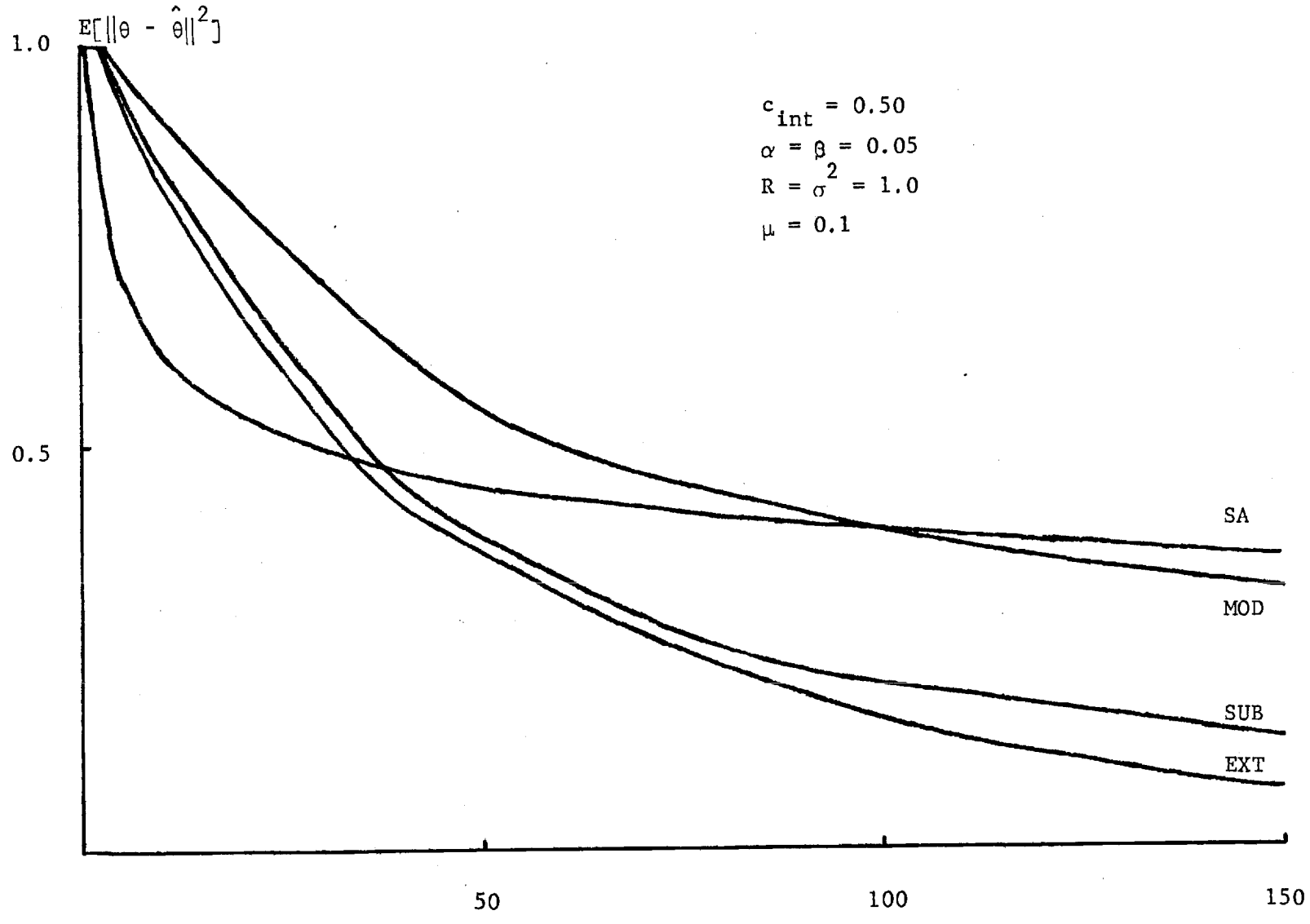


Figure 18. Comparison of Updating Algorithms.

III. A RE-EXAMINATION OF WALD's COSH TEST

For the application considered in this report, Wald's cosh test is unrealistic since the a priori probability assignment is constrained to assigning non-zero probability to only three values $\sigma\delta$, 0, $-\sigma\delta$. In fact, the mean γ has zero probability of ever assuming any one of these three values. For small values of $\sigma\delta$ any large deviation of the sample gradient mean from zero will also be far from $\pm\ \sigma\delta$. Thus, neither hypothesis is to be favored. Hence, a large number of observations will be taken before a decision can be made.

As a first step in developing a more realistic sequential test consider assigning the a priori probabilities

$$H_o : P_r\{\gamma = 0\} = 1$$

$$H_1 : p(\gamma) = \frac{1}{\sqrt{2\pi}\,\lambda} \exp\left[-\frac{\gamma^2}{2\lambda^2}\right] .$$

Choose the parameter λ such that

$$\sigma\delta = \int_{-\infty}^{\infty} \frac{|\gamma|}{\sqrt{2\pi}\,\lambda} \exp\left(-\frac{\gamma^2}{2\lambda^2}\right) d\gamma .$$

Hence,

$$\lambda = \sigma\delta \sqrt{\frac{\pi}{2}} .$$

The corresponding likelihood test is given by

$$\lambda_n = \sqrt{\frac{1}{1 + n\delta^2\pi/2}} \exp\left[\frac{\left(\frac{1}{n} S_n\right)^2}{2\sigma^2} \quad \frac{n^2\delta^2\pi/2}{1 + n\delta^2\pi/2}\right] ,$$

where

$$S_n = \sum_{i=1}^{n} Y_i .$$

The SPRT is given by: Continue taking observations as long as

$$b_n < S_n^2 < a_n ,$$

stop taking observations and decide to accept H_1 as soon as

$$S_n^2 \geq a_n ,$$

and stop taking observations and decide to accept H_o as soon as

$$S_n^2 \leq b_n ,$$

where

$$a_n = \sigma^2 \frac{1 + n\delta^2\pi/2}{\delta^2\pi/2} \ell n \left[A^2(1 + n\delta^2\pi/2\right]$$

and

$$b_n = \sigma^2 \frac{1 + n\delta^2\pi/2}{\delta^2\pi/2} \ell n \left[B^2(1 + n\delta^2\pi/2)\right] .$$

As shown by Cornfield [42] this test can be obtained from Wald's cosh test by using the weighting function

$$\omega(\theta) = \frac{1}{\sqrt{2\pi}\lambda} \exp\left(-\theta^2/2\lambda\right) .$$

The effect of the weighting function is to lower the stopping boundaries. This effect is somewhat similar to increasing the value of α in the original cosh test. However, one should have a little more control over the true error probabilities due to this modification.

A set of computer simulations similar to those for the other stopping rules were conducted for the above modification. The results are shown in Figure 19. Not unlike the previous results, the best performance is obtained with a large value of α and a small value of β.

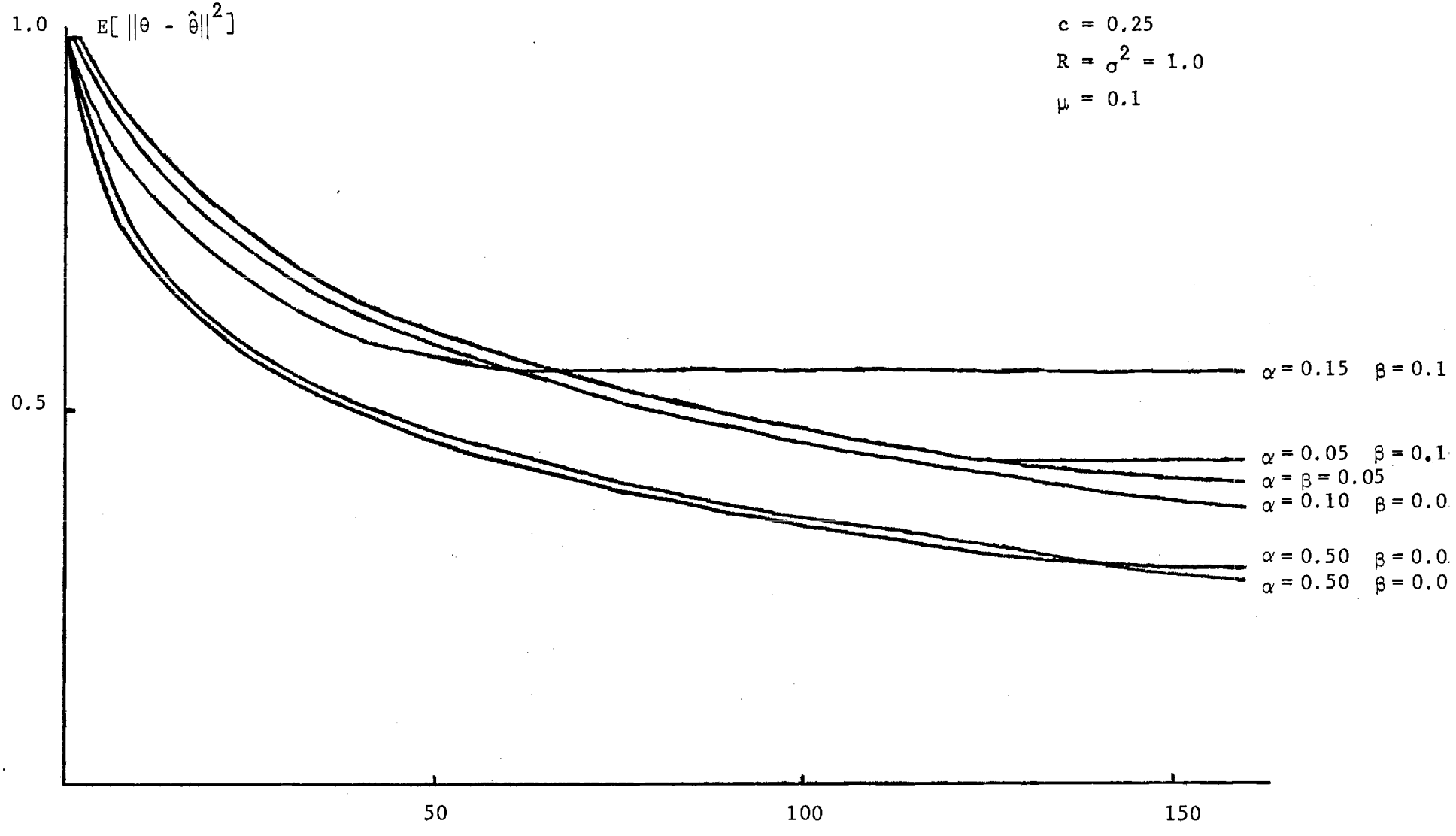


Figure 19. Learning Curves for Bayesian Cosh Test.

For a number of applications, such as the coding example to be considered in the next chapter, it is desirable to be able to guarantee termination after a finite predetermined number of observations. One method of doing this is the approach used by Chein and Fu [41]: Let $g_1(n)$ or $g_2(n)$ be either constants or monotonically nonincreasing and nondecreasing functions of n, respectively. Continue taking observations as long as

$$e^{g_2(n)} < \lambda_n < e^{g_1(n)} ,$$

stop taking observations and decide to accept H_1 as soon as

$$\lambda_n \geq e^{g_1(n)} ,$$

and stop taking observations and decide to accept H_o as soon as

$$\lambda_n \leq e^{g_2(n)} .$$

Within this framework, Wald's cosh test becomes: Continue taking observations as long as

$$g_2(n) + n\delta^2/2 > \ell n \cosh S_n < g_1(n) + n\delta^2/2 ,$$

stop taking observations and decide to accept H_1 as soon as

$$\ell n \cosh S_n > g_1(n) + n\delta^2/2 ,$$

and stop taking observations and decide to accept H_o as soon as

$$\ell n \cosh S_n < g_2(n) + n\delta^2/2 ,$$

where

$$S_n = \frac{\delta}{\sigma} \sum_{i=1}^{n} Y_i .$$

The modified cosh test is obtained in a straightforward fashion as done previously. Replace $\ell n \cosh S_n$ by $|S_n|$, $g_2(n)$ by $g_2(n) + \ell n\ 2$, and $g_1(n)$ by $g_1(n) + \ell n\ 2$.

The modification of the test discussed previously becomes: Continue taking observations as long as

$$b_n < S_n^2 < a_n ,$$

stop taking observations and decide to accept H_1 as soon as

$$S_n^2 \geq a_n ,$$

and stop taking observations and decide to accept H_o as soon as

$$S_n^2 \leq b_n ,$$

where

$$a_n = \frac{\sigma^2(1 + n\delta^2\pi/2)}{\delta^2\pi/2} \left[2g_1(n) + \ell n(1 + n\delta^2\pi/2)\right]$$

$$b_n = \frac{\sigma^2(1 + n\delta^2\pi/2)}{\delta^2\pi/2} \left[2g_2(n) + \ell n(1 + n\delta^2\pi/2)\right]$$

and

$$S_n = \sum_{i=1}^{n} Y_i .$$

These tests will be considered in the next chapter in connection with a coding problem. Note, that the separation between the boundaries is proportional to

$$\left[g_1(n) - g_2(n)\right]$$

Thus, the test terminates for n = N, where

$$g_1(N) = g_2(N) .$$

IV. CODING FOR ADDITIVE NOISE CHANNELS WITH FEEDBACK

Certain communication system problems may be modeled as follows: The channel consists of an additive white Gaussian noise forward link and a noiseless feedback link. In space exploration, for example, instruments periodically transmit their measurements back to Earth through a white Gaussian noise link. The link back to the satellite from Earth may be considered as a noiseless feedback link since the ground station has a very relaxed power constraint compared to the forward link.

Schalkwijk and Kailath [30] have discovered a coding scheme for this type of channel model that exploits the feedback to achieve considerable reductions in coding and decoding complexity over that needed for comparable performance using simplex codes for a one-way channel. The scheme is based on the Robbins-Monro stochastic approximation technique. Omura [32] has extended their coding scheme to the problem of transmitting an analog signal.

The basic communication system may be described as follows: Denote the message by θ. Define $M(X) = \frac{1}{\mu}(X - \theta)$, $\mu > 0$. Transmit a predetermined number $X = X_1$, known to the receiver, by transmitting the waveform

$$S(t) = M(X_1)\,\varphi(t) ,$$

where $\varphi(t)$ has support only on the interval $[0,T]$. The received signal,

$$r(t) = M(X_1)\,\varphi(t) + n(t) , \qquad 0 \le t \le T$$

is passed through the matched filter

$$h(t) = \varphi(T - t)$$

and the output sampled at $t = T$ to yield the statistic

$$Y_1 = M(X_1) + Z_1$$

where

$$Z_1 = \int_o^T n(t)\ \varphi(t)\ dt$$

and

$$\int_o^T \varphi^2(t)\ dt = 1$$

The receiver computes

$$X_2 = X_1 - \mu Y_1$$

and transmits this signal back to the satellite. Assuming no noise in the feedback loop, the satellite receives X_2 exactly. The satellite then transmits the signal

$$S(t) = M(X_2)\ \varphi(t - T)\ .$$

The receiver computes

$$X_3 = X_2 - \left(\frac{\mu}{2}\right) Y_2,$$

where

$$Y_2 = M(X_2) + Z_2$$

and

$$Z_2 = \int_T^{2T} n(t)\ \varphi(t - T)\ dt\ ,$$

and transmits this signal back to the satellite. In general, the satellite transmits the signal

$$s(t) = M(X_n)\ \varphi(t - nT + T)$$

and the receiver computes

$$X_{n+1} = X_n - \left(\frac{\mu}{n}\right) Y_n\ ,$$

where

$$Y_n = M(X_n) + Z_n$$

and

$$Z_n = \int_{(n-1)T}^{nT} n(t)\, \varphi(t + T - nT)\, dt ,$$

and transmits this signal back to the satellite.

The number of transmissions for a given message θ is predetermined by choosing an appropriate distortion criterion. For example, in the digital communication example considered by Schalkwijh and Kailath, the message θ is one of M possible values

$$\frac{1}{2M} , \frac{3}{2M} , \ldots , \frac{2M-1}{2M} .$$

The initial value for X is

$$X_1 = \frac{1}{2} .$$

It can be shown that for large n, the estimate X_n is normally distributed with mean θ and variance $N_o/2\alpha^2 n$, where $N_o/2$ is the (two-sided) spectral density of the white Gaussian noise. The probability of X_n not lying in the interval $\left[\theta - \frac{1}{2M} , \theta + \frac{1}{2M}\right]$ is the error probability

$$P_e = \text{erfc}_* \left(\frac{\sqrt{n}/\mu}{2M \sqrt{N_o/2}} \right)$$

where

$$\text{erfc}_*(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt .$$

Therefore, to guarantee that the error probability is less than or equal to some predtermined value ϵ, the number of transmission N is chosen to satisfy

$$\text{erfc}_* \left(\frac{1}{\mu M} \sqrt{\frac{N}{2N_o}} \right) \leq \epsilon/2 .$$

For a predetermined number of transmissions N and the transmission rate relative to channel capacity, $\gamma = R/C$, the optimum value of μ is

$$\mu = \sqrt{\frac{\gamma}{6N_o}}$$

The corresponding probability of error is

$$P_e = 2 \text{ erfc}_* \left(\sqrt{\frac{3N}{M^2}}\right) .$$

The relationship between γ, M, and N is given by

$$\gamma = \frac{2 \ell n M - 1}{\sum_{i=1}^{N-1} 1/j} .$$

Figures 20-21 present the error probability as function of N and M for various values of γ. (For $\gamma \geq 0.6$, refer to [30] for the corresponding curves.)

Note that this digital communication problem corresponds to the hypothesis testing problem

$$H_1 : |X_n - \theta| > \frac{1}{2M}$$

versus

$$H_o : |X_n - \theta| \leq \frac{1}{2M} .$$

Hence, the optional stopping stochastic approximation algorithm is applicable to this problem with

$$\beta = P_e ,$$

$$\sigma = N_o/2 = 1 ,$$

$$R = 1/\mu ,$$

and

$$\delta = 1/\mu M N_o = 1/2\mu M .$$

(This corresponds to $C = 1/8\mu M^2$.)

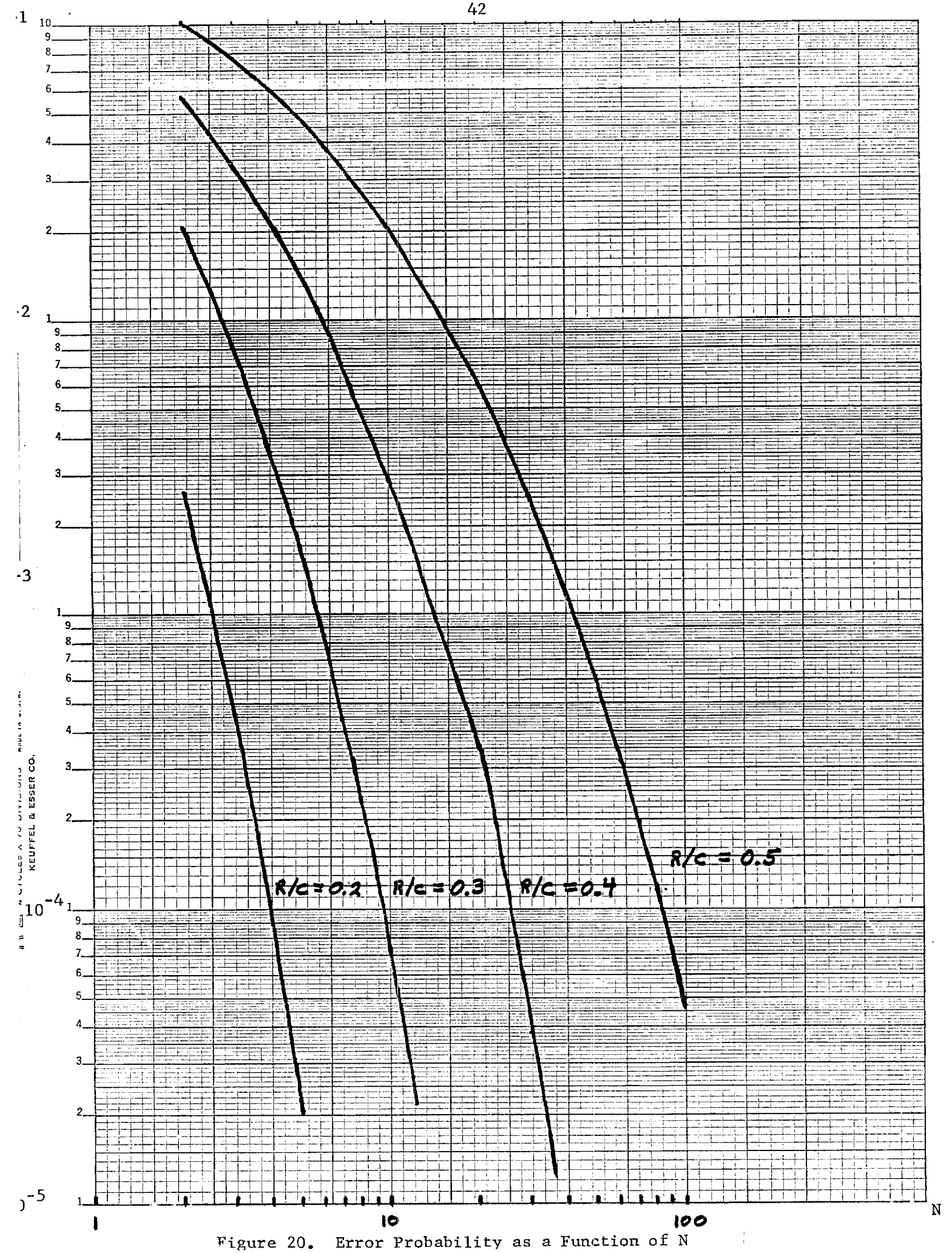


Figure 20. Error Probability as a Function of N

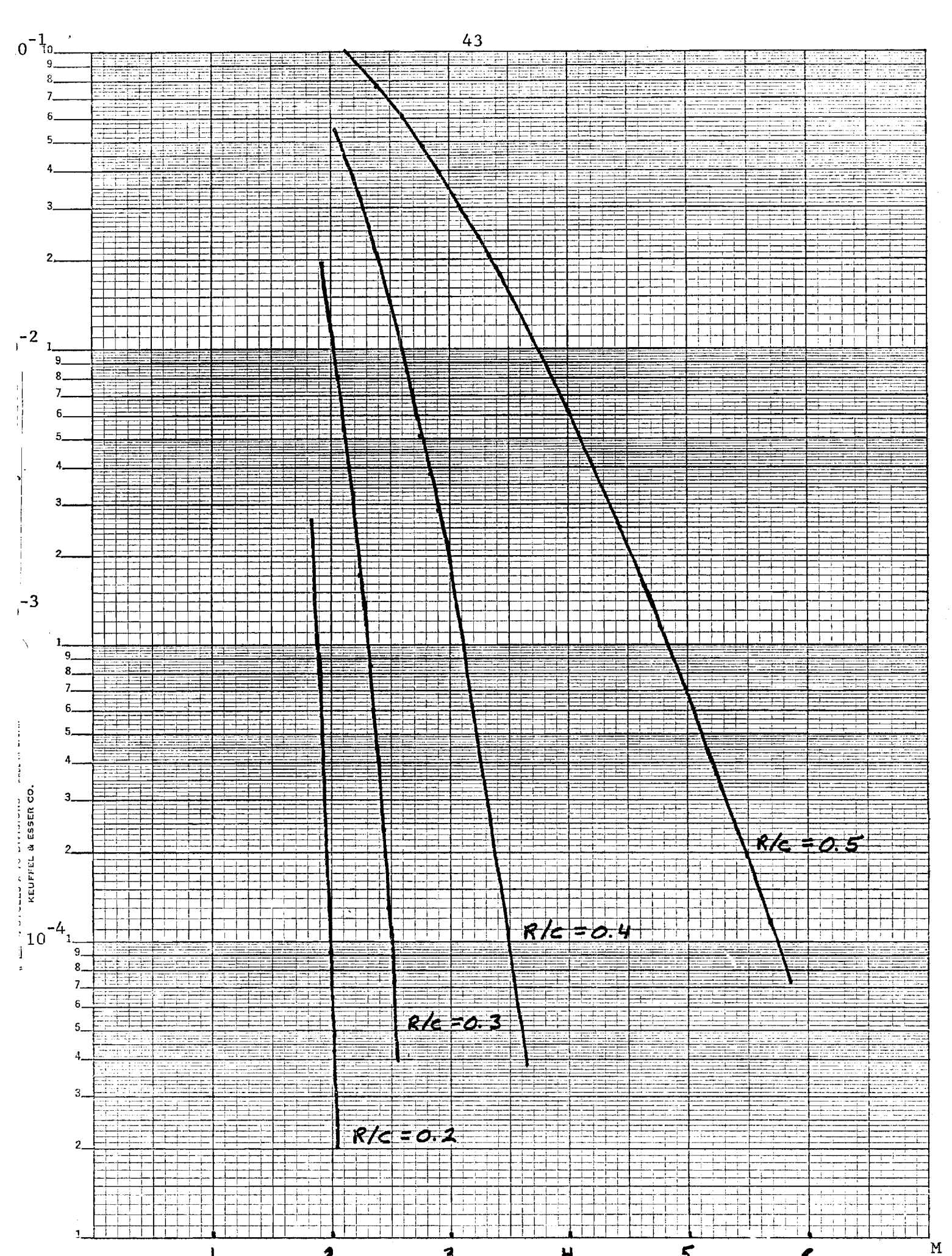


Figure 21 Error Probability as a Function of M

The Bayesian modification of Wald' cosh test was simulated for the case M = 4, N = 20. This corresponds to R/C = 0.493 with $P_e = 0.0058$. In order to achieve this error probability using the sequential test, the average number of transmission required was much greater than that required by stochastic approximation. This is not too surprising since the convergence rate for this algorithm is slower than for stochastic approximation theory for small values of N. The comparison for large N was prohibitive because of the number of trials that would be required.

In order to avoid the lengthy tests, the terminating boundary modification was next considered. The function $g_o(n)$ and $g_1(n)$ were chosen to be of the form

$$g_1(n) = \log A \left(1 - \frac{n}{N}\right)^r$$

and

$$g_o(n) = \log B \left(1 - \frac{n}{N}\right)^r ,$$

where A and B were chosen as before and $0 < r$. The prespecified termination time N corresponds to the predetermined number of transmissions for the Kailath-Schalkwijk coding scheme.

The Baysian modified cosh test was simulated with the terminating boundaries where the functions $g_1(n)$ and $g_2(n)$ were chosen to be of the form

$$g_1(n) = a' \left(1 - \frac{n}{N}\right)^{r_1}$$

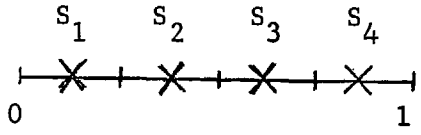
and

$$g_2(n) = b' \left(1 - \frac{n}{N}\right)^{r_2}$$

where $0 < r_1, r_2 \leq 1$, $a' > 0$, $b' > 0$, and N is the prespecified termination time. (For the modified cosh test using Wald's boundaries, $r_1 = r_2 = 0$, $a' = \log A$, and $b' = \log B$.) The results of those simulations are shown in

Table I for M = 4 with a desired error probability of $\beta = 0.0058$. Note that more than 20 iterations were required per decision because the decision H_1 was made occasionally, thus reintializing the boundaries. The blank entries were not simulated.

A similar simulation was performed for the modified cosh test using the terminating boundaries with the same parameters as the Bayesian modified test used. The results of those simulations are shown in Table II. An extensive study of this algorithm was not made since the performance on the simulations run was inferior to the Bayesian modified cosh test performance.



Message S_2 (or S_3) Transmitted

r \ α	β/2	β	2β	0.5	1-3β
0.5	0.00727	0.00685	0.00719	0.0107	0.0361
1.0	0.0189	0.0066	0.0102	0.0032	0.0624
2.0	0.0224	0.0288	0.0178	0.0284	—

Error Probability

r \ α	β/2	β	2β	0.5	1-3β
0.5	36.4	34.3	36.3	35.6	24.0
1.0	31.5	33.0	33.9	32.0	19.5
2.0	24.9	26.2	25.4	25.8	—

Average Number of Observations

Message S_1 (or S_4) Transmitted

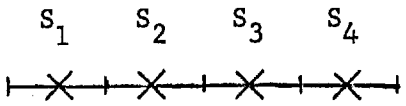
r \ α	β/2	β	2β	0.5	1-3β
0.5	—	0.0000	—	0.0108	—
1.0	—	0.00326	—	0.0132	—
2.0	—	0.0278	—	0.0260	—

Error Probability

r \ α	β/2	β	2β	0.5	1-3β
0.5	—	35.1	—	35.8	—
1.0	—	32.6	—	33.0	—
2.0	—	25.2	—	23.6	—

Average Number of Observations

Table I. Simulation Results for Bayesian Cosh Test (Terminating Boundaries, $\beta = 0.0058$)



Message S_2 (or S_3) Transmitted

r \ α	β/2	β	2β	0.5	1-3β
0.5	—	0.0167	—	—	—
1.0	0.0231	0.0138	0.289	0.0217	0.1684
2.0	—	0.0263	—	—	—

Error Probability

r \ α	β/2	β	2β	0.5	1-3β
0.5	—	33.4	—	—	—
1.0	28.9	27.8	28.9	27.2	8.9
2.0	—	23.9	—	—	—

Average Number of Observations

Table II. Simulation Results for Modified Cosh Test (Terminating Boundaries, $\beta = 0.0058$)

Based on these results, it is clear that the optional stopping stochastic approximation algorithm does not offer a method of improving the Schalkwick coding scheme for small values of N. A similar conclusion can be made for very large values of N since the coding scheme is shown to be optimal in this range [30]. For intermediate values of N, the algorithm may be of some use. However, due to the additional complexity of implementation, the algorithm is unlikely to provide a reasonable alternative.

One stopping rule that has not been considered in this research is the one proposed by Roberts and Mullis [44]. The optimal stopping boundaries are determined by using dynamic programming techniques when the maximum number of measurements are predetermined. The boundaries are more complex than those considered here. The additional improvement probably will not justify the additional complexity.

V. ANALYSIS OF THE MODIFIED COSH TEST

A complete theoretical analysis of the behavior of the Wald modified cosh test has not been achieved. However, the following theoretical results have been developed as a starting point.

Recall that the modified cosh test is to continue taking observations as long as

$$\ell n\ 2B + n\delta^2/2 < |S_n| < \ell n\ 2A + n\delta^2/2 ,$$

stop and decide H_1 as soon as

$$|S_n| \geq \ell n\ 2A + n\delta^2/2 ,$$

and stop and decide

$$|S_n| \leq \ell n\ 2B + n\delta^2/2 ,$$

where

$$A = (1 - \beta)/\alpha ,$$

$$B = \beta/(1 - \alpha) ,$$

and

$$S_n = \frac{\delta}{\sigma} \sum_{i=1}^{n} Y_i .$$

The random variables $\{Y_i\}$ are independent, identically distributed normally with mean

$$\mu = R(\hat{\theta} - \theta)$$

and variance σ^2. The behavior of this test can be assumed to be similar to the sequential test which takes a predetermined number of measurements

$$n_* = -\frac{2}{\delta^2} \log 2B$$

before applying the modified cosh sequential procedure [37] (See Figure 4). After n_*, the sequential test is equivalent to either the sequential test

$$\ell n\ 2B + n\delta^2/2 < S_n < \ell n\ 2A + n\delta^2/2$$

or

$$-\left[\ell n\ 2A + n\delta^2/2\right] < S_n < -\left[\ell n\ 2B + n\delta^2/2\right]$$

depending on whether $S_{n_*} > 0$ or $S_{n_*} < 0$, respectively.

Because of the symmetry, consider the details of the sequential test for $S_{n_*} > 0$ without loss of generality. This test corresponds to the sequential test for a sequence of independent, identically distributed random variables $\{Y_i\}$ according to the normal probability density function with variance σ^2 and mean

$$H_o : \mu = 0$$

$$H_1 : \mu = \sigma\delta$$

The true density of Y_i is normal with the mean μ and variance σ^2. The behavior of the modified cosh test may be expected to be accurately described by the operating characteristic function (OCF)

$$\mathcal{L}_T(\mu) = \mathcal{L}(\mu) + \mathcal{L}_{wc}(\mu)\ ,$$

where $\mathcal{L}(\mu)$ is the OCF of the above test and $\mathcal{L}_{wc}(\mu)$ is the probability that the wrong channel is entered $\left(S_{n_*} < 0\right)$. The probability density function for the number of observations for the modified cosh test conditional on μ and the terminal decision can be found from the conditional probability density function for the above test since

$$n_T = n_* + n$$

where n are the number of observations needed for this test. The stopping boundaries for this test are

$$\text{upper:}\quad -S_{n_*} + \ell n(A/B) + n\delta^2/2$$

$$\text{lower:}\quad -S_{n_*}$$

Note that the stopping boundaries are random, depending on the value of S_{n_*}.

Conditional on S_{n_*}, it has been shown [40] that

$$\mathcal{L}(\mu, S_{n_*}) = \frac{C^h - 1}{C^h - D^h}$$

where

$$\ell n\, C \triangleq - S_{n_*} + \ell n(A/B) ,$$

$$\ell n\, D \triangleq - S_{n_*} ,$$

and

$$h = 1 - \frac{2\mu}{\sigma\delta} .$$

Moreover, the characteristic function for n conditional on the terminal decision is given by [40]

$$M_o(w, \mu, S_{n_*}) \triangleq E\left[e^{jwn} \mid \text{acc } H_o, \mu, S_{n_*}\right]$$

$$= \frac{C^{w_2(w)} - C^{w_1(w)}}{\mathcal{L}(\mu, S_{n_*}) \left[D^{w_1(w)} C^{w_2(w)} - D^{w_2(w)} C^{w_1(w)}\right]}$$

and

$$M_1(w, \mu, S_{n_*}) \triangleq \left[e^{jwn} \mid \text{acc } H_1, \mu, S_{n_*}\right]$$

$$= \frac{D^{w_1(w)} - D^{w_2(w)}}{\left[1 - \mathcal{L}(\mu, S_{n_*})\right]\left[D^{w_1(w)} C^{w_2(w)} - D^{w_2(w)} C^{w_1(w)}\right]}$$

where

$$w_1(w) = -\frac{\mu}{\sigma\delta} + \sqrt{\left(\frac{\mu}{\sigma\delta}\right)^2 - 2jw/\delta^2}$$

and

$$w_2(w) = -\frac{\mu}{\sigma\delta} - \sqrt{\left(\frac{\mu}{\sigma\delta}\right)^2 - 2jw/\delta^2}$$

In order to determine the OCF and the characteristic function for n as required, one must average the above expressions over S_{n_*} given that the correct channel is selected; i.e. the above test is, in fact, the sequential test used. Since S_{n_*} is normally distributed with mean $n_*\mu\delta/\sigma$ and variance $n_*\delta^2$, it is straightforward to establish that

$$p_{S_{n_*}}(x|\mu) = \begin{cases} \dfrac{1}{1 - \mathcal{L}_{wc}(\mu)} \dfrac{\exp\left[-\dfrac{\left(x + \frac{2\mu \ln 2B}{\sigma\delta}\right)^2}{4\ln(1/2B)}\right]}{\sqrt{4\pi \ln(1/2B)}} & 0 \le x \le \ln \frac{A}{B} \\ 0 & \text{otherwise} \end{cases}$$

where

$$\mathcal{L}_{wc}(\mu) = \text{erfc}_*\left[\frac{\mu}{\sigma\delta}\sqrt{2\ln(1/2B)}\right] - \text{erfc}_*\left[\sqrt{n_*}\left(\frac{\mu}{\sigma} + \frac{\delta}{2}\right) + \frac{\ln 2A}{\sqrt{n_*}\,\delta}\right]$$

$$\cong \text{erfc}_*\left[\frac{\mu}{\sigma\delta}\sqrt{2\ln(1/2B)}\right]$$

and

$$\text{erfc}_*(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-t^2/2\right) dt \ .$$

The characteristic function for n conditional on the terminal decision and the value of μ given by

$$M_i(w,\mu) = \int M_i\left(w,\mu,S_{n_*}\right) p_{S_{n_*}}\left(x|\,\mu, \text{acc } H_i\right) dx$$

By Bayes rule,

$$p_{S_{n_*}}(x|\mu, \text{acc } H_i) = \begin{cases} \dfrac{\mathcal{L}(\mu,x)\; p_{S_{n_*}}(x|\mu)}{\mathcal{L}(\mu)} & i = 0 \\ \dfrac{[1 - \mathcal{L}(\mu,x)]\; p_{S_{n_*}}(x|\mu)}{1 - \mathcal{L}(\mu)} & i = 1 \end{cases}$$

Hence,

$$M_o(W,\mu) = \frac{1}{\mathcal{L}(\mu)} \int \frac{\left[C^{W_2(W)} - C^{W_1(W)}\right] p_{S_{n_*}}(x|\mu)\, dx}{\left[D^{W_1(W)} C^{W_2(W)} - D^{W_2(W)} C^{W_1(W)}\right]}$$

$$M_1(W,\mu) = \frac{1}{[1 - \mathcal{L}(\mu)]} \int \frac{\left[D^{W_1(W)} - D^{W_2(W)}\right] p_{S_{n_*}}(x|\mu)\, dx}{\left[D^{W_1(W)} C^{W_2(W)} - D^{W_2(W)} C^{W_1(W)}\right]}$$

and

$$\mathcal{L}(\mu) = \int \mathcal{L}(\mu,x)\; p_{S_{n_*}}(x|\mu)\, dx.$$

Carrying out the integration for $\mathcal{L}(\mu)$, one obtains

$$\mathcal{L}(\mu) \cong \frac{\text{erfc}_*\left[-\left(1+\frac{\mu}{\sigma\delta}\right)\sqrt{2\ell n(1/2B)}\right]}{1 - \mathcal{L}_{wc}(\mu)} - \exp\left[-\left(\frac{2\mu}{\sigma\delta} + 1\right)\ell n\left(\frac{1}{2B}\right)\right]$$

$$\cong \frac{\text{erfc}_*\left[-\left(1+\frac{\mu}{\sigma\delta}\right)\sqrt{2\ell n(1/2B)}\right]}{1 - \text{erfc}_*\left[\frac{\mu}{\sigma\delta}\sqrt{2\ell n(1/2B)}\right]} \exp\left[-\left(\frac{2\mu}{\sigma\delta} + 1\right)\ell n\left(\frac{1}{2B}\right)\right]$$

for

$$\ell n\left(\frac{A}{B}\right) \gg -2\left(1 + \frac{\mu}{\delta}\right)\ell n(1/2B)\ .$$

Similar laborious calculations are possible for M_o and M_1. However, for

large A/B the terms

$$\frac{-C^{W_1(W)}}{-D^{W_2(W)} C^{W_1(W)}} = D^{-W_2(W)} = e^{xw_2(w)}$$

and

$$\frac{-D^{W_2(W)}}{-D^{W_2(W)} C^{W_1(W)}} = C^{-W_1(W)} = \left(\frac{A}{B}\right)^{W_1(W)} e^{xw_1(w)}$$

dominate. Hence,

$$M_o(W,\mu) \cong \frac{\text{erfc}_* \left[\sqrt{2\left[\left(\frac{\mu}{\sigma\delta}\right) - 2jW/\delta^2\right] \ell n(1/2B)}\right]}{\mathcal{L}(\mu)} \cdot \exp\left[W_1(W)\, \ell n(1/2B)\right]$$

and

$$M_1(W,\mu) \cong \frac{\text{erfc}_* \left[-\sqrt{2\left[\left(\frac{\mu}{\sigma\delta}\right)^2 - 2jW/\delta^2\right] \ell n(1/2B)}\right]}{1 - \mathcal{L}(\mu)} \cdot \exp\left[W_2(W)\, \ell n(1/2B)\right]$$

The analysis has not been carried any further than this. Note that, in principle, one may obtain the complete characterization of the optional stopping stochastic approximation algorithm from these experessions since they relate the properties of the stopping rule to the value of $\mu = R(\hat{\theta} - \theta)$.

There is one interesting observation that one may make. For A/B large, the theoretical results suggest that for a given value of R, C, and σ, algorithms with the same value of B should behave similarly. The simulation shown in Figures 8, 10, and 12 correspond to B = 0.0527, 0.0555, and 0.0588

and A = 19, 9.5, and 6.3, respectively. Note the similarity of the "SUB" curves.

VI. REFERENCES

1. A. Wald, Sequential Analysis, John Wiley & Sons, Inc., New York, 1947.

2. H. Robbins and S. Monro, "A Stochastic Approximation Methods," Ann. Math. Stat., 22, 1951, pp. 400-407.

3. J. R. Blum, "Multidimensional Stochastic Approximation Methods," Ann. Math. Stat., 25, 1954, pp. 737-744.

4. A. Dvoretzky, "On Stochastic Approximation," Proc. 3rd Berkeley Symp. on Math. Stat. and Prob., J. Neyman (Ed.), vol. 1, University of California Press, Berkeley California, 1956.

5. M. T. Wasan, Stochastic Approximation, Cambridge University Press, London, 1969.

6. Y. T. Chien and K. S. Fu, "On Bayesian Learning and Stochastic Approximation," IEEE Trans. Systems Science and Cybernetics, SSC-3, June 1967, pp. 28-38.

7. Y. T. Chien and K. S. Fu, "Learning in Non-Stationary Environment Using Dynamic Stochastic Approximation," Proc. 5th Allerton Conf. on Circuit and System Theory, 1967, pp. 337-345.

8. D. B. Cooper, "Adaptive Pattern Recognition and Signal Detection Using Stochastic Approximation," IEEE Trans. Electronic Computers EC-13, June 1964, pp. 306-307.

9. R. DeFigueiredo, "Convergent Algorithms for Pattern Recognition in Non-Linearly Evolving Nonstationary Environment," Proc. IEEE, 56, February 1968, pp. 188-189.

10. J. M. Schumpert and S. S. Yan, "Stochastic Approximation Nonparametric Training Procedures for Multi-Category Pattern Classifiers," Proc. 5th Allerton Conf. on Circuit and System Theory, 1967, pp. 792-799.

11. R. L. Kashyap and C. C. Blaydon, "Recovery of Functions from Noisy Measurements Taken at Randomly Selected Points and Its Applications to Pattern Classification," Proc. IEEE, 54, August 1966, pp. 1127-1129.

12. Y. C. Ho and R. C. K. Lee, "Identification of Linear Dynamic Systems," Information and Control, 8, February 1965, pp. 93-100.

13. H. J. Kushner, "A Simple Iterative Procedure for the Identification of the Unknown Parameters of a Linear Time-Varying Discrete System," Trans. ASME, J. Basic Engrg., 85 ser. D, June 1963, pp. 227-235.

14. D. J. Sakrison, "The Use of Stochastic Approximation to Solve the System Identification Problem," IEEE Trans. Automatic Control, AC-12, October 1967, pp. 563-567.

15. G. N. Saridis and G. Stein, "Stochastic Approximation Algorithms for Linear Discrete-Time System Identification," IEEE Trans. Automatic Control, AC-13, October 1968, pp. 515-523.

16. G. N. Saridis and G. Stein, "A New Algorithm for Linear System Identification," IEEE Trans. Automatic Control, AC-13, October 1968, pp. 592-594.

17. G. N. Saridis, K. S. Fu, and Z. J. Nikolic, "Stochastic Approximation Algorithms for System Identification, Estimation, and Decomposition of Mixtures," IEEE Trans. Syst. Sci. and Cybernetics, SSC-5, January 1969, pp. 8-15.

18. J. M. Mendel, "Gradient Identification for Linear Systems," in Adaptive, Learning, and Pattern Recognition Systems, J. M. Mendel and K. S. Fu (eds.), Academic Press, New York, 1970.

19. J. K. Holmes, "Two Stochastic Approximation Procedures for Identifying Linear Systems," IEEE Trans. Automatic Control, AC-14, June 1969, pp. 292-295.

20. J. P. Comer, "Some Stochastic Approximation Procedures for Use in Process Control," Ann. Math. Stat., 35, 1964, pp. 1136-1145.

21. Y. C. Ho and B. H. Whalen, "An Approach to the Identification and Control of Linear Dynamic Systems with Unknown Parameters," IEEE Trans. Automatic Control, AC-8, July 1963, pp. 255-256.

22. J. Skalnsky, "Learning Systems for Automatic Control," IEEE Trans. Automatic Control, AC-11, January 1966, pp. 6-19.

23. G. E. P. Box and G. M. Jenkins, "Some Statistical Aspects of Adaptive Optimization and Control," J. Roy. Stat. Soc. Ser. B, 24, pp. 297-343.

24. Y. C. Ho, "On Stochastic Approximation and Optimum Filtering Methods," J. Math. Anal. Appl., 6, No. 1, February 1963, pp. 152-154.

25. D. J. Sakrison, Application of Stochastic Approximation Methods to Optimum Filter Design," IRE International Convention Record, 9, part 4, 1961.

26. L. D. Davisson, "A Theory of Adaptive Filtering," IEEE Trans. Info. Theory, IT-12, April 1966, pp. 97-102.

27. L. A. Gardner, Jr. "Adaptive Predictors," Trans. 3rd Prague Conf. on Info. Theory, Stat. Dec. Functions, and Random Processes, 1960, pp. 123-192.

28. R. W. Lucky, "Equalization of Digital Communication Systems," B.S.T.J., 45, February 1966, pp. 255-286.

29. L. J. Griffiths, "A Simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays," Proc. IEEE, 57, October 1969, pp. 1696-1704.

30. J. P. N. Schalkwijk and T. Kailath, "A Coding Scheme for Additive Noise Channels with Feedback - Part I: No Bandwidth Constraint," IEEE Trans. Info. Thy., IT-12, April 1966, pp. 172-182.

31. J. P. N. Schalkwijk, "A Coding Scheme for Additive Noise Channels with Feedback - Part II: Band-Limited Signals," IEEE Trans. Info. Thy., IT-12, April 1966, pp. 183-195.

32. J. K. Omura, "Optimum Linear Transmission of Analog Data for Channels with Feedback," IEEE Trans. Info. Thy., IT-14, Jan. 1968, pp. 38-43.

33. T. Kailath, "An Application of Shannon's Rate-Distortion Theory to Analog Communication Over Feedback Channels," Proc. IEEE (Letters) 55, June 1967, pp. 1102-1103.

34. C. E. Shannon, "Coding Theorems for a Discrete Source with Fidelity Criterion," in Decision Processes, R. E. Machol, Ed., New York: McGraw-Hill, 1963.

35. M. Sobel and A. Wald, "A Sequential Decision Procedure for Choosing One of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution," Ann. Math. Stat., vol. 20, 1949, pp. 502-522.

36. P. Armitage, "Sequential Analysis with More than Two Alternative Hypotheses, and Its Relation to Discriminant Function Analysis," Suppl. J. Roy. Statist. Soc., vol. 12, 1950, pp. 137-144.

37. G. D. Simons, "A Sequential Three Hypothesis Test for Determining the Mean of a Normal Population with Known Variance," Ann. Math. Stat., vol. 28, 1967, pp. 1365-1375.

38. T. W. Anderson, "A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size," Ann. Math. Stat., vol. 21, 1960, pp. 165-197.

39. G. B. Wetherill, Sequential Methods in Statistics, Methuen: London, 1966.

40. J. V. DiFranco and W. L. Rubin, Radar Detection, Prentice Hall, Englewood Cliffs, 1968 (Chapter 16).

41. J. J. Bussgang and D. Middleton, "Optimum Sequential Detection of Signal in Noise," IRE Trans. Inform. Theory, vol. IT-1, Dec. 1955, pp. 5-18.

42. M. Hecht and M. Schwarz, "M-ary Sequential Detection for Amplitude-Modulated Signals in One or Two Dimensions," IEEE Trans. Commun. Technol., vol. COM-16, Oct. 1968, pp. 669-675.

43. S. Nishikawa, "Sequential M-ary PAM System," IEEE Trans. Commun. Technol., vol. COM-21, Jan. 1973, pp. 22-33.

44. R. A. Roberts and C. T. Mullis, "A Bayes Sequential Test of M Hypotheses," IEEE Trans. Inform. Theory (Corresp.), vol. IT-16, Jan. 1970, pp. 91-94.

45. L. C. Palmer, "Sequential Tests to Select among M Hypotheses," IEEE Trans. Inform. Theory (Corresp.), vol. IT-18, Jan. 1972, pp. 211-214.

46. G. Simons, "Lower Bounds for Average Sample Number of Sequential Multi-Hypothesis Tests," Ann. Math. Stat., vol. 38, 1968, pp. 1343-1364.

47. W. Hoeffding, "Lower Bounds for the Expected Sample Size and the Average Risk of a Sequential Procedure," Ann. Math. Stat., vol. 31, 1960, pp. 352-368.

48. M. Skibinsky, "Some Properties of a Class of Bayes Two-Stage Tests," Ann. Math. Stat., vol. 31, 1960, pp. 332-351.

49. G. Schwarz, "Asymptotic Shapes of Bayes Sequential Testing Regions," Ann. Math. Stat., vol. 33, 1962, pp. 224-236.

50. J. Cornfield, "A Bayesian Test of Some Classical Hypotheses -- With Applications to Sequential Clinical Tests," Journal of American Statistical Assoc., vol. 61, Sept. 1966, pp. 577-594.

51. Y. T. Chien and K. S. Fu, "A modified sequential recognition machine using time-varying stopping boundaries," IEEE Trans. Info. Theory, vol. IT-12, 1966, pp. 206-214.

VII. ADDITIONAL INFORMATION

A. Personnel

Dr. Brown, as principal investigator, was the only personnel involved in this research activity.

B. Publications

No papers have been prepared or are in preparation as of the date of submission of this final report. It is anticipated that one paper based on this research will be prepared.

C. Other Activities

No other activities were conducted related to this grant.